# CHARACTERIZATION OF DEFENSE-ASSOCIATED GENE EXPRESSION IN *PHYTOLACCA AMERICANA* USING NOVEL TRANSCRIPTOMIC AND GENOMIC RESOURCES

KIRA NELLER

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
**DOCTOR OF PHILOSOPHY**

GRADUATE PROGRAM IN BIOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

August 2019

**Abstract**

*Phytolacca americana* (pokeweed) is a non-model plant with agriculturally relevant characteristics, being a heavy metal hyperaccumulator resistant to virus infection. Pokeweed synthesizes pokeweed antiviral protein (PAP), an RNA glycosidase thought to function in defense. Study of this plant, and of the biological relevance of PAP, has been hindered by a lack of reference sequences. Here, the first mRNA transcriptome, miRNA transcriptome, and de novo genome assembly of pokeweed are reported. The mRNA transcriptome was initially built using plants subjected to jasmonic acid (JA), a hormone that mediates resistance to pathogens. Using this resource, ~59,000 transcripts and ~52,000 natural antisense transcript (NAT) pairs were annotated. Differential expression analysis revealed that many transcripts were JA-responsive, including several PAP isoforms and PAP-specific NAT pairs. The annotated mRNA transcriptome was also used to investigate miRNA control of JA response. Novel miRNA/target pairs involved in hormone biosynthesis, signal transduction, and pathogen defense were predicted. One of these comprised a transcript encoding a JA biosynthesis enzyme, which was validated, representing the first reported miRNA-induced cleavage event of a non-transcription factor target of the JA-response pathway. For approximately half of the pairs, expression of the miRNA and target was positively correlated, suggesting a biologically relevant regulatory dynamic. To gain additional knowledge of the pokeweed defense response, a comprehensive mRNA transcriptome was obtained from plants subjected to different stresses. This transcriptome was used to annotate the first pokeweed de novo genome assembly, revealing the structures of ~30,000 genes. Several known PAP isoform genes were identified, as well as a putative novel isoform and PAP pseudogenes. The gene model of each PAP isoform contained a long intron in the 5' untranslated region, and preliminary evidence suggests it regulates gene expression. Differential expression

analysis confirmed that in addition to JA, several PAP isoforms responded to other stresses. PAP isoforms were co-expressed with genes involved in terpenoid biosynthesis, JA-mediated signalling, and metabolism of amino acids and carbohydrates. Taken together, this study generated the first large-scale sequence references for pokeweed and provided insight into the regulation of defense-associated gene expression in this non-model plant.

**Acknowledgements**

Thank you to my supervisor, Dr. Kathi Hudak, for your guidance, encouragement, and trust in me as we ventured into a new research direction. For your care to recognize and nurture the unique talents of your students, creating an environment where we thrive together and individually. For long and meandering chats about science or not. For editing with meticulous attention to detail.

Thank you to my family for endless love, support, and wisdom. For faithfully sitting through practice presentations. For being interested in my work and what it means. To my dad, for always bringing me home. To my mom, for reminding me to do the things that scare me. To my sister, for showing me what true strength is.

Thank you to my friends in the Hudak lab. To Alex, for always smiling, making me laugh, and fixing everything. To Camille, for your graphic design magic, eye for all things aesthetic, and for hanging out with me at the house after seeing me all day. To Jen, for your convictions, your endless trivia, and our long walks.

Thank you to my friends outside of the lab. For reminding me that there is life outside of the lab.

**Table of contents**

## List of tables

### Chapter 2



### Chapter 3



### Chapter 4

## List of figures

### Chapter 1



### Chapter 2



### Chapter 3



### Chapter 4

**List of abbreviations**

| | |
|---|---|
| ABA | Abscisic acid |
| AED | Annotation edit distance |
| AGO | Argonaute |
| AP2/ERF | Apetala2/ethylene-responsive factor |
| bZIP | Basic leucine zipper domain |
| COI1 | Coronatine-insensitive 1 |
| CRE | Cis-regulatory element |
| DCL1 | Dicer-like 1 |
| DEG | Differentially expressed gene |
| E or ET | Ethanol |
| FC | Fold change |
| FDR | False discovery rate |
| FPKM | Fragments per kilobase per transcript per million mapped reads |
| GO | Gene ontology |
| GUS | β-glucuronidase |
| JA | Jasmonic acid |
| JA-Ile | (+)-7-iso-JAIle |
| JAZ | Jasmonate zim domain |
| miRNA | Micro RNA |
| MYC | Myelocytomatosis |
| NAT | Natural antisense transcript |
| NPR | Non-expressor of PR Genes |
| ORF | Open reading frame |
| PAP | Pokeweed antiviral protein |
| PCC | Pearson correlation coefficient |
| PDF1.2 | Plant defensin 1.2 |
| PEG | Polyethylene glycol |
| PP2C | 2C protein phosphatase |
| PR | Pathogenesis-related protein |
| pre-miRNA | miRNA precursor |
| pri-miRNA | primary miRNA |
| PYR/PYL/RCAR | Pyrabactin resistance/pyrabactin resistance–like/regulatory components of ABA receptor |
| qRT-PCR | Quantitative reverse transcription PCR |
| RACE | Rapid amplification of cDNA ends |
| RIP | Ribosome-inactivating protein |
| RPM | Reads per million |

| | |
|---|---|
| rRNA | Ribosomal RNA |
| SA | Salicylic acid |
| siRNA | Short interfering RNA |
| SnRK2 | SNF1-related protein kinase |
| TGA | TGACG sequence-specific binding protein |
| TMM | Trimmed mean of m-values |
| TPM | Transcripts per million |
| TSS | Transcription start site |
| UTR | Untranslated region |
| VSP2 | Vegetative storage protein 2 |
| WND | Wounded with forceps |
| WT | Watered normally |

# CHAPTER 1:

# Introduction

### 1.1 Pokeweed and its relevance in medicine and agriculture

*Phytolacca americana*, commonly referred to as pokeweed, belongs to the Phytolaccaceae family of flowering plants comprising 65 species of herbs, shrubs, and trees. Pokeweed is native to eastern North America and parts of the USA (Midwest, Gulf Coast, and west coast), with naturalization in Europe, the West Indies, and Asia. This non-model plant synthesizes metabolites and proteins that have been studied for their medical and agricultural applications. Among them is pokeweed antiviral protein (PAP), a type of ribosome-inactivating protein (RIP; rRNA N-glycosidase; EC 3.2.2.22). In a process termed depurination, RIPs remove a specific adenine from the conserved α-sarcin/ricin loop of large rRNA, thereby arresting protein synthesis in vitro (Endo et al., 1988). The resulting toxicity has led to investigation of some RIPs as immunotoxins for cancer therapy (Zeng et al., 2015). In addition, PAP has broad-spectrum activity against plant and animal viruses when expressed in heterologous systems (Dai et al., 2003; Ishag et al., 2013; Mansouri et al., 2009; Wang et al., 1998; Zhabokritsky et al., 2014; Zoubenko et al., 1997, 2000). In both its native pokeweed and transgenic plants, PAP is localized outside the cell in the apoplasm (Ready et al., 1986). It is hypothesized that PAP gains entry to the cytoplasm during pathogen infection, at which point ribosome inactivation inhibits the synthesis of host and viral proteins. In addition to rRNA, PAP depurinates the genomes of some RNA viruses, interfering directly with viral replication and packaging (He et al., 2008; Karran and Hudak, 2008; Lodge et al., 1993; Mansouri et al., 2009; Rajamohan et al., 1999; Zhabokritsky et al., 2014). Other products isolated from pokeweed include terpenoids with anti-inflammatory and antimicrobial properties (Woo and Kang, 1976) and pokeweed mitogen, used as a laboratory reagent to induce proliferation of lymphocytes (Bekeredjian-Ding et al., 2012). Pokeweed has also attracted interest for its potential in phytoremediation. As a hyperaccumulator, the plant is capable of growing in soil containing

high levels of metals, absorbing and concentrating them in its tissues (Liu et al., 2010; Peng et al., 2008; Zhao et al., 2011).

### 1.2 Ribosome inactivating protein (RIP) and pokeweed antiviral protein (PAP) isoforms

RIPs are synthesized mainly by plants; they are highly represented in certain families but narrowly distributed overall, given their presence in ~20% of angiosperm taxonomic orders (Di Maro et al., 2014). Plant RIPs are generally encoded as isoforms by multi-gene families (Chan et al., 2011; Urasaki et al., 2017; Wu et al., 2015; Wytynck et al., 2017). Although the biological relevance of RIP isoforms in plants is not completely understood, particularly with some being restricted temporally or spatially, numerous reports of induced expression upon diverse stresses suggest a role in defense. RIP expression is upregulated in response to pathogen infection, insect feeding, cold/heat, drought, salinity, mechanical wounding, or stress-associated phytohormones (Iglesias et al., 2005; Jiang et al., 2008; Qin et al., 2009; Reinbothe et al., 1994; Song et al., 2000; Tartarini et al., 2010). Several isoforms of PAP have been reported in pokeweed. PAP-I (Irvin, 1975), PAP-II (Irvin et al., 1980), and PAP-III (Rajamohan et al., 1999) are leaf isoforms synthesized in spring, early, and late summer, respectively. Isoforms have also been purified from seeds (PAP-S/S1 and PAP-S2; Barbieri et al., 1982; Honjo et al., 2002) and roots (PAP-R; Bolognesi et al., 1990), while PAP-α (Kataoka et al., 1992) was found to be expressed throughout the plant, and PAP-H (Park et al., 2002) and PAP-C (Barbieri et al., 1989) were identified in cultured cells. PAP isoforms vary in their ability to depurinate different templates (Honjo et al., 2002; Rajamohan et al., 1999). However, little is known about the regulation and roles of discrete PAP isoforms, nor that of RIPs in general. In pokeweed, this question is further complicated by

the high sequence similarity of PAP isoforms, which makes them difficult to distinguish through standard methods such as qRT-PCR and Western blotting.

### 1.3 Hormones implicated in plant defense

Given that the expression of many RIPs is responsive to changes in plant hormone levels, what follows is a brief overview of those hormones most cited as affecting RIPs. **Figure 1** summarizes the hormone signalling pathways and their key nodes of crosstalk.

#### 1.3.1 Jasmonic acid (JA)

JA is a lipid-derived molecule synthesized from α-linolenic acid, the latter found in plastid membranes. It plays a major role in defense against necrotrophic pathogens and insect herbivores; it also regulates developmental transitions in the plant (Wasternack and Song, 2017). The biologically active form of the hormone is an isoleucine conjugate, (+)-7-iso-JAIle (JA-Ile) (Fonseca et al., 2009), which is perceived through a co-receptor formed by the F-box protein Coronatine-insensitive 1 (COI1) and a member of the Jasmonate ZIM domain (JAZ) family (Chini et al., 2007; Thines et al., 2007). JAZ proteins are negative regulators of transcription; in the absence of JA-Ile, they bind and inhibit members of the MYC and AP2/ERF transcription factor (TF) families (Chini et al., 2007; Thines et al., 2007). Increasing levels of JA-Ile promote the interaction of JAZ and COI1, resulting in ubiquitination and degradation of the JAZ co-repressor, freeing TFs to activate downstream gene expression (Chini et al., 2007; Thines et al., 2007). An activated JA defense response is associated with induced expression of marker genes VSP2 (an acid phosphatase with insecticidal activity), PDF1.2 (a defensin), and basic pathogenesis-related (PR) proteins (Ali et al., 2018; Wasternack and Song, 2017). The JA signalling pathway is divided

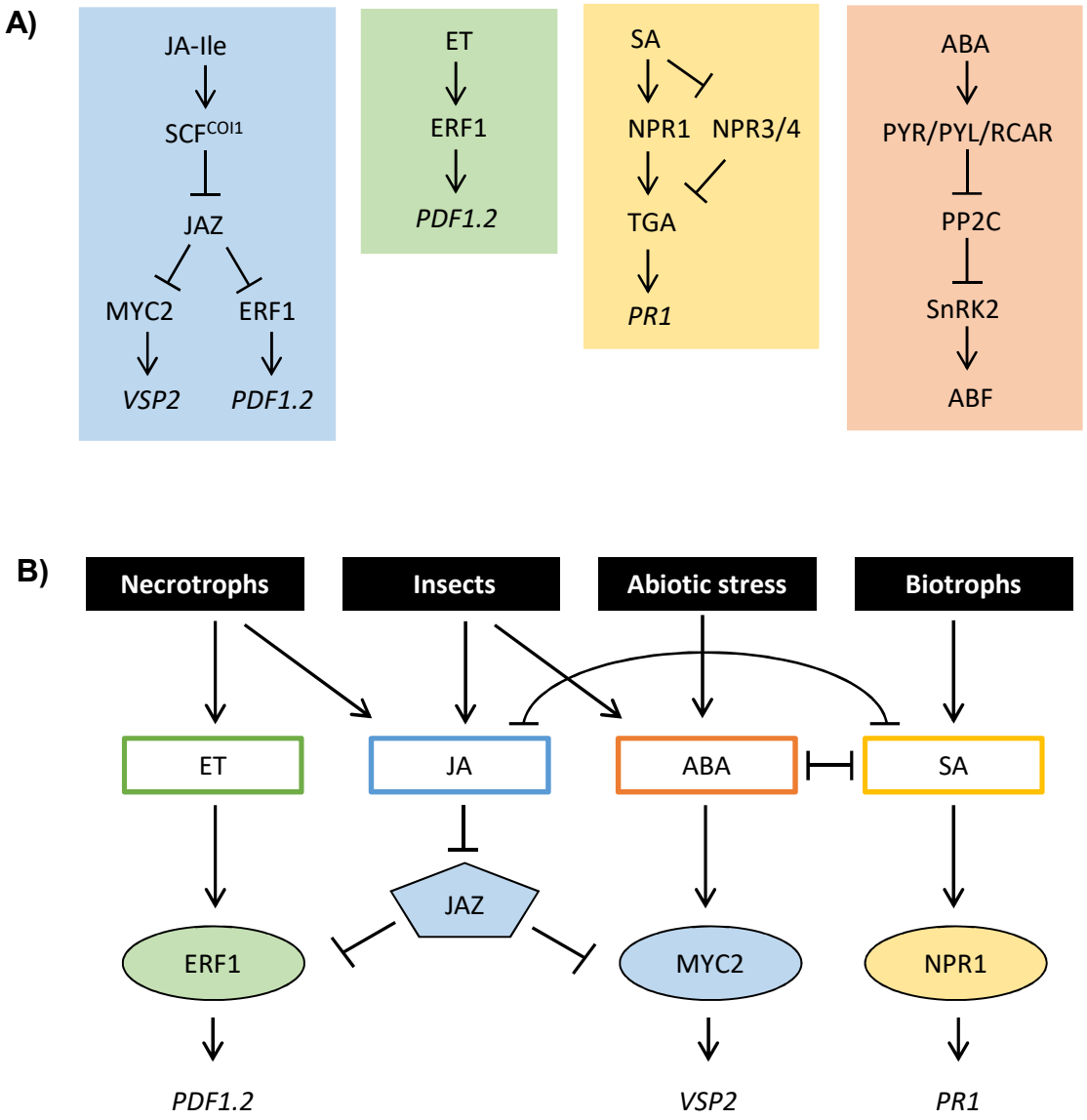



**Figure 1. Plant hormones involved in defense. A)** Summary of individual hormone signalling pathways for jasmonic acid (JA), ethylene (ET), salicylic acid (SA), and abscisic acid (ABA). **B)** Key nodes of hormone crosstalk in response to pathogens, insects, and abiotic stress. Individual pathway colours correspond to those in A). Transcription factors are represented by circles. Inhibitory interactions are indicated by a line ending with a single tick. Pathways that are mutually antagonistic are indicted by a line with ticks on both ends. Marker genes associated with each pathway are shown.

into two mutually antagonistic branches regulated by MYC and ERF TFs (Lorenzo et al., 2004). The MYC branch is associated with wound response and defense against insect herbivores, while the ERF branch mediates resistance to necrotrophic pathogens. Insect attack specifically activates the MYC branch and results in high expression of VSP2, while necrotrophs induce expression of PDF1.2 through the ERF branch. Ethylene acts synergistically with JA via the ERF branch. Both ethylene and JA are required for induced systemic resistance, an enhanced state of plant defense arising from root colonization of beneficial microbes (Pieterse et al., 1998).

### 1.3.2 Salicylic acid (SA)

SA is important for plant innate immunity against biotrophic pathogens and is required for systemic acquired resistance (Backer et al., 2019). Biosynthesis of SA occurs through the cinnamic acid pathway (Dempsey et al., 2012) and the isochorismate pathway (Wildermuth et al., 2001), the latter route predominant during pathogen infection. Elucidation of SA receptors has involved some disagreement, as both NPR1 (Non-expressor of PR Genes 1; Wu et al., 2012) and NPR3/4 (Fu et al., 2012) have been proposed. NPR1 is a transcriptional co-activator for members of the TGA family (Despres et al., 2007). NPR1 exists in two forms in the cell (Tada et al., 2008). When SA levels are low, NPR1 is sequestered in the cytoplasm as an oligomer due to the presence of intermolecular disulphide bonds. When SA levels increase during pathogen infection, redox changes in the cytosol enable NPR1 monomerization, at which point it enters the nucleus and promotes gene expression. Only recently was it determined that both NPR1 and NPR3/4 are SA receptors, but with opposite roles (Ding et al., 2018). NPR3/4 are repressors of TGA TFs, binding them independently of NPR1. In the absence of SA, NPR3/4 bind to and repress these TFs; presence of SA abolishes this interaction, enabling NPR1 to bind instead and activate downstream

gene expression. This includes upregulation of acidic PR1, a marker gene of systemic acquired resistance (Ali et al., 2018). Although both SA and JA mediate pathogen defense, their signalling pathways tend to be mutually antagonistic. SA antagonism of JA occurs through inhibition of JA biosynthesis via cytosolic NPR1 (Spoel et al., 2003) and TGA TFs (Zander et al., 2010), while JA-mediated antagonism of SA signalling involves MYC2 (Laurie-Berry et al., 2006).

### 1.3.3 Abscisic acid (ABA)

ABA mediates tolerance to environmental stresses including drought, salinity, cold, and heat, mainly by regulating stomatal opening; it is also important in seed development, shoot elongation, root growth, and establishing plant dormancy (Tarkowská and Strnad, 2018). As with JA, synthesis of the isoprenoid-derived ABA originates in plastids. ABA is perceived by plasma and nuclear membrane proteins of the PYR/PYL/RCAR (Pyrabactin Resistance/Pyrabactin resistance–Like/Regulatory Components of ABA Receptor) family (Ma et al., 2009; Park et al., 2009). In the absence of ABA, type 2C protein phosphatases (PP2Cs) bind to and inhibit the action of SNF1-related protein kinases (SnRK2s), preventing them from activating downstream targets (Fujii et al., 2005; Park et al., 2009). In the presence of ABA, recruitment of PP2C by PYR/PYL/RCAR leads to dissociation of SnRK2 and PP2C. This enables SnRK2 to phosphorylate TFs of the ABA-Responsive Element-Binding Factor (ABF) family (Furihata et al., 2006; Uno et al., 2000) and ion channels (Lee et al., 2013). In terms of crosstalk, ABA exerts its effects by modulating the JA and SA signalling pathways (Ku et al., 2018). It acts in combination with JA to activate MYC2 (Anderson et al., 2004) and negatively regulates SA signalling by reducing levels of SA biosynthesis enzymes (Audenaert et al., 2002). The relationship between ABA and systemic

acquired resistance is mutually antagonistic (Yasuda et al., 2008), indicating trade-off between abiotic and biotic stress responses.

### 1.4 Plant small RNAs and miRNA biogenesis

In addition to hormones, small RNAs (~20-24 nt) are important regulators of gene expression in plants (Lee and Carroll, 2018). Small RNAs fine-tune gene expression by binding to target RNAs, resulting in transcriptional or post-transcriptional gene silencing. They are divided into two general classes: microRNAs (miRNAs) and short interfering RNAs (siRNAs). miRNAs are derived from single-stranded transcripts that adopt an imperfect stem-loop structure, and siRNAs from perfect, double-stranded RNA precursors generated by RNA-dependent RNA polymerases. siRNAs are further classified into subtypes that vary in origin, mechanism of action, required components, and level of sequence conservation across plants. However, biogenesis of miRNAs is more homogenous (Yu et al., 2017), and the mature sequences show greater levels of species conservation (Baldrich et al., 2018). As shown in **Figure 2**, miRNA biosynthesis begins with transcription of a *MIR* gene by RNA polymerase II, producing a primary miRNA (pri-miRNA) that consists of a stem-loop region flanked by unstructured arms. In sequential steps, dicer-like 1 RNase (DCL1) excises the stem-loop to form the miRNA precursor (pre-miRNA), then generates a duplex (~21 nt) comprising the miRNA and its opposing strand, classically termed miRNA* (Fukudome and Fukuhara, 2017). Excision of the duplex from the stem-loop is based on structural features, including the length of base-paired stem proximal and distal to the miRNA/miRNA*, and presence of a terminal loop. Following methylation (Li et al., 2005), the miRNA is loaded into an Argonaute (AGO) protein, typically AGO1, establishing the RNA-induced silencing complex. Until recently, miRNA loading was thought to occur only in the

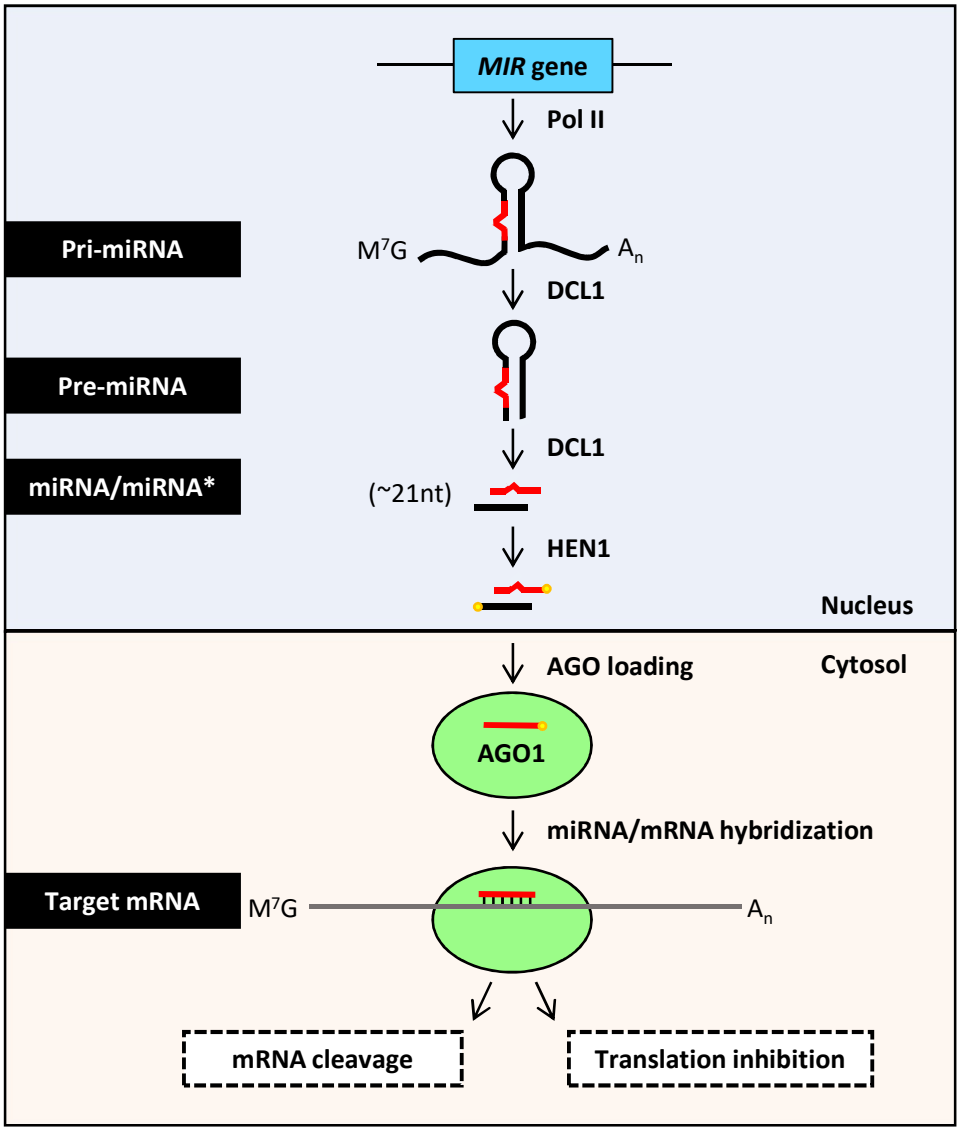



**Figure 2. miRNA biogenesis and function.** A *MIR* gene is transcribed by Pol II to yield a non-coding, single-stranded transcript that folds back on itself, forming a bulged hairpin flanked by unstructured arms. This primary miRNA (pri-miRNA) is processed mainly by DCL1 to a miRNA precursor (pre-miRNA) and miRNA/miRNA* duplex (~21nt) in sequential steps. The duplex is 3'-end methylated by HEN1 to protect from degradation, then exported to the cytosol. The miRNA guide strand is selected and incorporated into the RNA-induced silencing complex (RISC), which contains an AGO protein, usually AGO1. The RISC binds to a target mRNA on the basis of sequence complementarity and either slices it or inhibits its translation. Figure from: Neller et al., 2019.

cytosol. However, new evidence supports a revised model in which loading can occur in the nucleus, followed by nucleo-cytoplasmic shuttling of the AGO1:miRNA complex (Bologna et al., 2018). Selection of the mature miRNA is not random, with dependence on miRNA precursor processing factors, the 5′-terminal nucleotide, and structure of the small RNA duplex (Yu et al., 2017). Plant miRNAs tend to have high sequence complementarity with targets and act primarily through mRNA cleavage, although translational inhibition has also been observed (Chen, 2004; Gandikota et al., 2007). Given that targets of plant miRNAs often encode transcription factors, they are poised to modulate diverse processes during plant growth, development, and stress response (Li et al., 2017).

### 1.5 Transcriptome and genome assembly

Genome-wide analyses are very useful and increasingly common to study patterns of gene expression under different conditions. Of course, these require genomic resources for the species under study, such as an annotated transcriptome or genome (Bolger et al., 2018). Transcriptome and genome sequencing are termed RNA-seq and whole genome sequencing, respectively. Different methods are in use that produce different read-lengths. Second-generation sequencing (e.g. Illumina) is widely used; it produces short reads (up to 250 bp) that must be assembled to reconstruct the original, full-length sequence. Third-generation sequencing (e.g. PacBio, Oxford Nanopore) is capable of providing much longer reads (up to 20 Kb), but it is cost-restrictive for most. Regardless of the chosen strategy, it is likely that some assembly will be required to obtain a eukaryotic transcriptome or genome sequence. As shown in **Figure 3**, reads can be assembled in a de novo (reference-free) or guided (reference-based) manner (Moreton et al., 2016). In the case of a newly studied species without a reference genome, de novo assembly is performed by

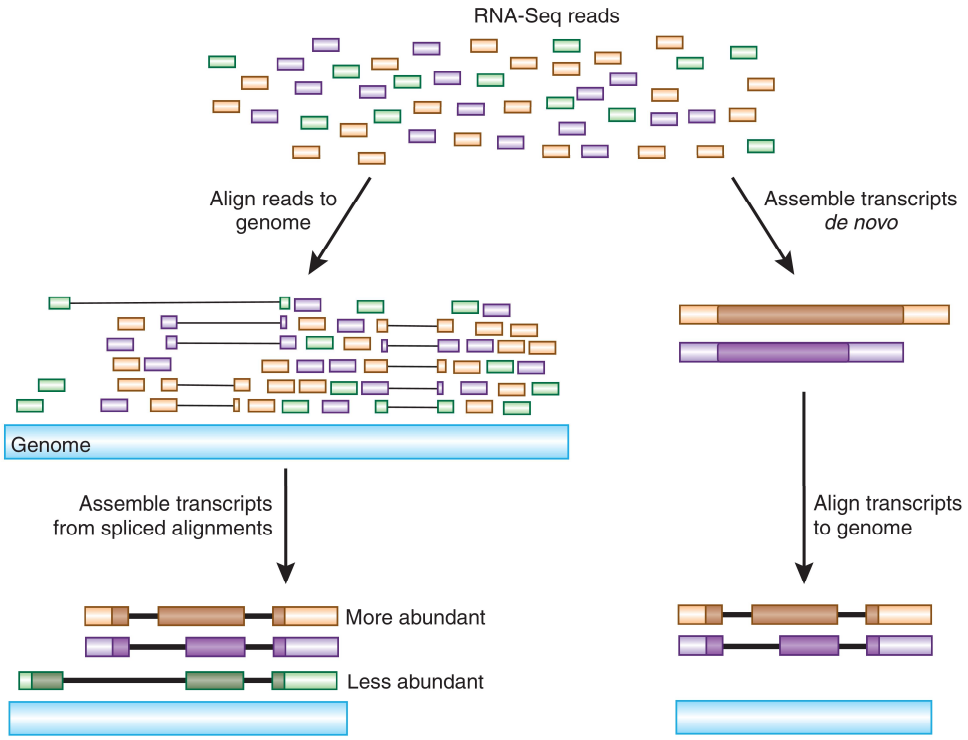



**Figure 3. Strategies for reconstructing and annotating transcripts from RNA-seq reads.** The genome-guided approach (left), first aligns short RNA-seq reads to a reference genome in a splice-aware manner, then reconstructs transcripts from the spliced alignments. The de novo assembly approach (right), assembles transcripts directly from the RNA-seq reads. To delineate exon/intron structures, these transcripts are aligned to the genome. A de novo assembly approach is likely to assemble only the most abundant transcripts. RNA-seq reads are coloured according to the transcript isoform from which they originated. Protein-coding regions, obtained during annotation, are indicated in darker colours. Figure from: Haas et al., 2010.

identifying overlapping sequences between reads and joining them to produce a contiguous sequence. Since this is a computationally intensive process, a de Bruijn graph is used to compact the representation of overlapping and non-overlapping sequences (**Figure 4**). Here, all subsequences of length k in reads are termed k-mers. A de Bruijn graph is constructed using unique k-mers as nodes, with edges indicating overlapping k-mers. That is, if a k-mer is shifted by one base (k-1) and it overlaps another k-mer, an edge is drawn between those nodes. A linear chain of k-mer nodes is compressed into one node when two nodes are joined by a single unique edge. By traversing all paths through the graph, each independent sequence can be reconstructed. For a transcriptome, this produces all transcript variants (i.e. those arising from alternative splicing and paralogs). When applied to a genome, this process identifies paralogs and allelic variants. It is important to note that a reference genome is a 'scaffold' assembly that traverses only the highest coverage path through each region, making downstream analysis more manageable (**Figure 5**). De novo assembly is a challenging computational problem. At the transcriptome level, it can be difficult to distinguish transcripts arising from alternative splicing or paralogous genes due to their shared k-mer sequences. As for plant genomes, they contain long transposons and repeat-rich regions that are difficult to assemble, and polyploidy greatly expands the de Bruijn graph (Kyriakidou et al., 2018). If a reference genome is available for the species under study or a highly related one, it greatly eases transcript reconstruction and genome assembly by orienting aligned reads.

Once a transcriptome or genome sequence is available, alignment of the short RNA-seq reads back to this reference provides a measure of gene expression level based on read counts (Finotello and Di Camillo, 2015). This process is repeated for all samples, often with the goal of identifying differentially expressed genes (DEGs) between conditions. A typical workflow

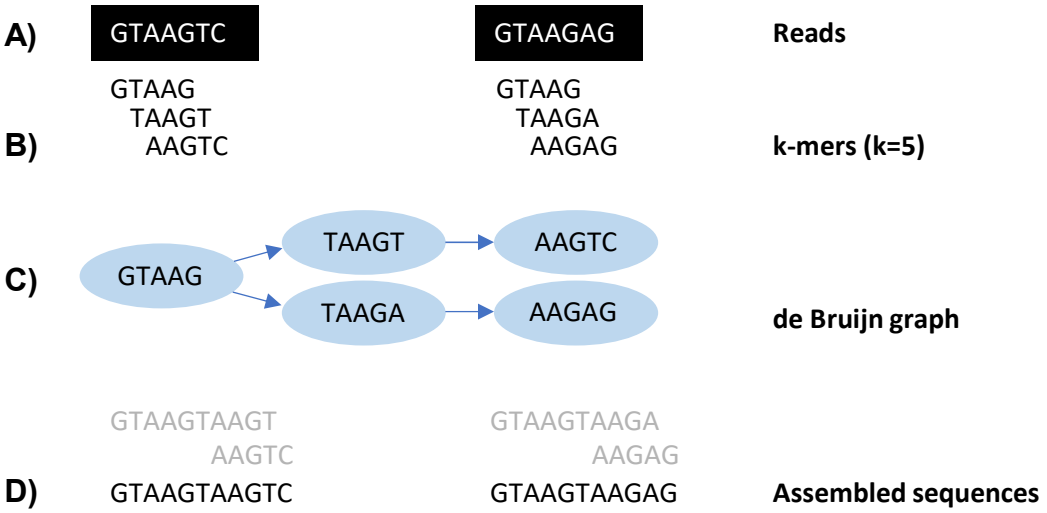



**Figure 4. A simple de Bruijn graph. A)** Read sequences. **B)** All possible k-mers of length 5 from the reads. **C)** A de Bruijn graph constructed from unique k-mers as nodes and overlapping k-mers connected by edges. A k-mer shifted by 1 base overlaps another k-mer by k-1 bases. **D)** Assembled sequences obtained by traversing the two possible paths in the graph. Figure adapted from: Moreton et al., 2016.

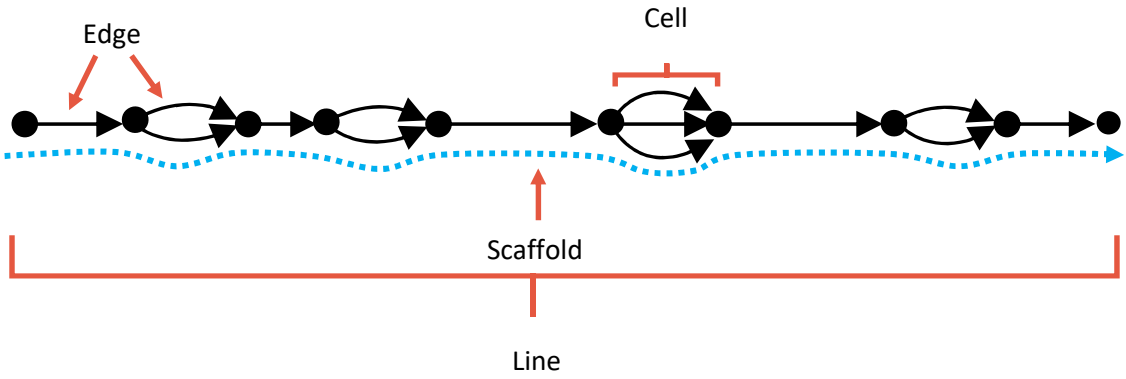



**Figure 5. Terminology associated with an assembly.** An assembly is a graph of all potential paths through a sequence region. An 'edge' is the largest continuous region that could be assembled for the region. A 'line' is a series of connected edges that could no longer be extended. A 'cell' is a region that contains more than one path. A 'scaffold' is the single, highest coverage path through a line; that is, the path with the greatest number of reads supporting it.

involves the following steps: (i) for each sample, reads are aligned to the reference sequence; (ii) the total number of reads aligning to each gene is counted; (iii); counts are normalized based on sequencing depth and gene length; (iv) counts of each gene are tested for significant difference between conditions. If a reference genome is unavailable, differential expression analysis can be performed using the de novo assembled transcriptome as a substitute. However, there may be inconsistencies between transcript-level and gene-level analyses, especially in the case of alternatively spliced genes (Song et al., 2019).

### 1.6 Genome annotation

Following genome assembly, genes are annotated on the basis of predicted gene models (Bolger et al., 2018). A gene model is a structural representation of an mRNA transcript that contains information such as coding sequence, exon/intron boundaries, and untranslated regions (UTRs). Due to alternative splicing, it is possible to have more than one model for a given gene. A standard genome annotation pipeline is as follows: (i) repetitive sequences (e.g. transposons) are masked to prevent their interference in subsequent steps; (ii) transcript and protein sequences from the species under study or related species are aligned to the genome; (iii) alignments are processed to infer exon/intron boundaries; (iv) ab initio gene prediction is performed; (v) results from alignments and ab initio predictions are used to construct a gene model with an associated quality score. Termed the annotation edit distance (AED), this measure quantifies how well the gene model agrees with the information from which it was derived. Different computational approaches have been implemented to select a final gene model among candidates, including winner-takes-all, consensus, or weighted consensus; (vi) conserved protein domains in genes are identified based on sequence comparison, and their associated gene ontology (GO) terms are

obtained. GO terms are structured, curated concepts relating to gene functions. A gene product may be annotated in up to three categories: molecular function, cellular component, and biological process. The ontology is organized as a directed graph in which each term has defined relationships to one or more other terms. Lower-level terms have greater specificity than higher-level ones.

Annotation generally comprises two sources of information: extrinsic 'evidence' from transcript and protein alignments, and intrinsic ab initio prediction. The latter involves searching the genomic DNA sequence for features of protein-coding genes, including specific sequences or statistical characteristics. Incorporation of ab initio prediction is desirable, but it requires availability of a trained model for the species under study. In the case of a novel genome annotation, this model can be obtained and refined in an iterative manner consisting of multiple rounds of gene prediction (Campbell et al., 2014a). In Round 1, gene models are predicted solely from transcript and protein evidence. Top gene models as per AED score are provided to the ab initio predictor to obtain a preliminary model. In Round 2, this preliminary model is used in combination with transcript/protein evidence to obtain a new set of gene structures. This is referred to as 'hint-based ab initio prediction', reflecting the use of evidence to guide ab initio prediction. Top gene models from Round 2 are provided to the predictor as before to refine the model, followed by a final Round 3, the output of which is generally considered a trustworthy set of gene models.

### 1.7 Bioinformatic prediction and annotation of miRNAs

Large-scale characterization of miRNAs requires integration of small RNA sequencing data and a reference genome or transcriptome for the species under study (Neller et al., 2019). A typical workflow includes miRNA prediction, identification of conserved and novel sequences,

and functional analysis. Computational miRNA prediction is based on experimentally determined features of miRNA biogenesis. The process is summarized as follows: (i) small RNAs are aligned to the reference genome/transcriptome; (ii) a portion of the reference sequence surrounding each alignment is extracted and folded, representing a candidate miRNA precursor; (iii) the likelihood that a particular small RNA sequence originated from the biological processing of a miRNA precursor is evaluated based on information from alignments and structural compatibility. Top-scoring miRNAs have detectable miRNA* reads and precursors that form stable hairpin structures. The miRNA* sequence is the passenger strand of the mature miRNA duplex; since it is often degraded, it is usually detected only for highly abundant miRNAs. Following prediction, miRNAs are annotated as conserved or novel by conducting sequence similarity searches against a database of known plant miRNAs. If a significant hit is identified in another plant species, the miRNA is deemed conserved; otherwise, it is novel. Most miRNAs are species-specific or conserved across a few closely related species (Baldrich et al., 2018). These miRNAs typically have low abundance, heterogeneous processing, and are encoded from single genes rather than multi-gene families. In contrast, miRNAs conserved across plants tend to be highly abundant. To obtain insight into the biological relevance of predicted miRNAs, functional annotation is performed. For each miRNA, putative targets are identified by searching the reference genome/transcriptome for binding sites, i.e. sequences that are complementary to the miRNA and amenable to base-pairing. Proteins encoded by RNA targets are then characterised based on their associated GO terms.

### 1.8 Goals of study

The study of pokeweed has been hindered by a lack of large-scale sequence datasets. Therefore, the aim of my PhD project was to develop the first genome and transcriptome references for this non-model plant. Using these bioinformatic resources, I investigated pokeweed gene expression under abiotic and biotic stress treatments. This work has provided insight into defense-associated gene expression in pokeweed, including the endogenous regulation and biological relevance of PAP.

This project is divided into three parts:

1. Assembly and annotation of the mRNA transcriptome and identification of JA-responsive mRNAs.

2. miRNA/target prediction and construction of an interaction network based on correlated expression in response to JA.

3. Assembly and annotation of the genome and identification of stress-responsive gene clusters.

# CHAPTER 2:

## The pokeweed leaf mRNA transcriptome and its regulation by jasmonic acid

Published manuscript

**Neller, K. C. M.**, Klenov, A., and Hudak, K. A. (2016). *Front. Plant Sci.* 7, 1–13. doi:10.3389/fpls.2016.00283.

**Contributions:** KN and KH conceived the study. KN performed all bioinformatic analyses. AK conducted plant treatments and wet-lab validations (qRT-PCR and cloning of novel sequences). KN drafted the manuscript. KH edited the manuscript.

All Supplementary Material referenced in this manuscript (**Supplementary Data Sheets 1-6**, **Supplementary Figures 1-4**) is available at the above doi.

## 2.1 Abstract

The American pokeweed plant, *Phytolacca americana*, is recognized for synthesizing pokeweed antiviral protein (PAP), a ribosome inactivating protein (RIP) that inhibits the replication of several plant and animal viruses. The plant is also a heavy metal accumulator with applications in soil remediation. However, little is known about pokeweed stress responses, as large-scale sequencing projects have not been performed for this species. Here, we sequenced the mRNA transcriptome of pokeweed in the presence and absence of jasmonic acid (JA), a hormone mediating plant defense. Trinity-based *de novo* assembly of mRNA from leaf tissue and BLASTx homology searches against public sequence databases resulted in the annotation of 59,096 transcripts. Differential expression analysis identified JA-responsive genes that may be involved in defense against pathogen infection and herbivory. We confirmed the existence of several PAP isoforms and cloned a potentially novel isoform of PAP. Expression analysis indicated that PAP isoforms are differentially responsive to JA, perhaps indicating specialized roles within the plant. Finally, we identified 52, 305 natural antisense transcript pairs, four of which comprised PAP isoforms, suggesting a novel form of RIP gene regulation. This transcriptome-wide study of a Phytolaccaceae family member provides a source of new genes that may be involved in stress tolerance in this plant. The sequences generated in our study have been deposited in the SRA database under project # SRP069141.

## 2.2 Introduction

The pokeweed plant, *Phytolacca americana,* is a member of the Phytolaccaceae family of flowering plants, which includes 65 species of herbs, shrubs and trees. Pokeweed is native to eastern North America and has become naturalized in Europe, the West Indies and Asia. This species is of interest because it synthesizes pokeweed antiviral protein (PAP), a ribosome inactivating protein (RIP) with RNA *N*-glycosidase activity. Several isoforms of PAP are reported to exist in pokeweed, exhibiting different temporal (PAP-I, PAP-II, PAP-III) and spatial (PAP-S, PAP-R, PAP-alpha) expression patterns (Barbieri et al., 1982; Bolognesi et al., 1990; Irvin and Kelly, 1980; Irvin, 1975; Kataoka et al., 1992; Rajamohan et al., 1999). RIPs are present in less than 20% of angiosperm taxonomic orders, and phylogenetic analysis indicates a complex evolutionary history (Di Maro et al., 2014). They are potent defense proteins effective against a range of viruses, fungi, and less commonly, insects (Stirpe, 2013).

Pokeweed has broad applications in agriculture and medicine. PAP inhibits the replication of several plant and animal viruses, either through ribosome inactivation which limits viral proliferation, or by direct depurination of the viral genome (He et al., 2008; Karran and Hudak, 2008; Lodge et al., 1993; Mansouri et al., 2009; Rajamohan et al., 1999). Interestingly, recent work demonstrates that pokeweed accumulates high levels of heavy metals, especially cadmium and manganese, with promising applications in soil detoxification (Dou et al., 2009; Zhao et al., 2011, 2012). Nevertheless, little is known about pokeweed, as the genes involved in stress response have not been identified.

Here, we have sequenced the pokeweed mRNA transcriptome in the presence and absence of jasmonic acid (JA) treatment. JA is a plant hormone that mediates defense against herbivores and necrotrophic pathogens. As herbivores are often viral vectors, the JA pathway also has

important implications for virus resistance. We showed recently that PAP mRNA and protein levels increase in the presence of JA (Klenov et al., 2015). A link between jasmonate and other plant RIPs has previously been established. For example, PIP2 from *Phytolacca insularis*, ME1 from *Mirabilis expansa*, and JIP60 from barley are induced by JA or its methyl jasmonate derivative (Dunaeva et al., 1999; Song et al., 2000; Vepachedu et al., 2003). Furthermore, expression of the insecticidal maize RIP2 is increased 100-fold at the RNA level upon caterpillar feeding, demonstrating the relevance of RIPs in anti-herbivory (Chuang et al., 2014). By sequencing the transcriptome of pokeweed treated with JA, we will gain novel information about the regulation of specific PAP isoforms and how these proteins are integrated within the larger network of pokeweed pathogen response. This work lays the important foundation to understand the resiliency of pokeweed to biotic and abiotic factors.

We report the *de novo* assembly and annotation of the pokeweed mRNA transcriptome from leaf tissue. Through a combination of differential expression and gene ontology (GO) analysis, we identified JA-responsive genes and enriched GO terms involved in stress and defense. We confirmed the existence of several published PAP isoforms, reported their distinct responses to JA, and cloned a potentially novel PAP isoform. Finally, we report the discovery of PAP natural antisense transcripts (NATs) that are also JA-responsive, which may represent a novel form of RIP gene regulation.

## 2.3 Material and methods

### Pokeweed growth conditions and jasmonic acid treatment

Pokeweed seeds were treated with 37% sulfuric acid for 5 minutes and submerged in water for 4 days at room temperature. Seeds were germinated in soil (Promix BX) and maintained in a

growth chamber (AC60, Biochambers, MN) under fluorescent and incandescent lights at 180 μE $m^{-2}$ $s^{-1}$ with periods of 16 h day and 8 h night. Fertilizer was provided once every 2 weeks with N:P:K 20:20:20. For experimental treatment, plants were sprayed with 5 mL of 5 mM JA dissolved in 0.5% ethanol (to improve the solubility of JA). Negative control plants were sprayed with 0.5% ethanol. Following treatment, plants were returned to the chamber and leaf tissue was harvested 24 hours later. All plants used in this study were at the 4-leaf stage of growth.

**Total RNA isolation, library construction and sequencing**

Total RNA was extracted from leaf tissue of pokeweed plants treated with 5 mM JA in 0.5% ethanol or 0.5% ethanol alone using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA). An equal amount of total RNA from three independent plants was pooled for each biological replicate. In total, six mRNA libraries were generated (n=3 per treatment, from 18 total plants). Libraries were constructed with the TruSeq Stranded mRNA Library Preparation Kit (RS-122-2101, Illumina). Sequencing was performed on a single lane of an Illumina HiSeq 2500 machine by The Centre for Applied Genomics (The Hospital for Sick Children, Toronto, Canada) to generate paired-end reads of 150 bases. Raw sequences are available at the SRA database under project # SRP069141.

**RNA-seq data processing and de novo transcriptome assembly**

Prior to assembly, adapters were clipped, low-quality bases were trimmed ($Q < 30$, averaged over four bases) and synchronicity of paired-end files was maintained using Trimmomatic v. 0.32.1 (Bolger et al., 2014) as follows: PE -phred33 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:30. The pokeweed transcriptome was assembled with

Trinity v. r2014-04-13p1 (Grabherr et al., 2011) using the following command-line, which invoked paired-end, stranded information: --seqType fq --JM 48G --left reads-1.fq --right reads-2.fq --SS_lib_type RF --CPU 24.

**Transcriptome annotation and refinement**

Trinotate v. 2.0.1 (Haas et al., 2013) was used for transcriptome annotation. BLAST (Altschul et al., 1997) searches were conducted against both the SwissProt (Bairoch and Boeckmann, 1991) and UniRef90 (Suzek et al., 2007) databases (current as of Jan, 2015). Owing to strand-specific sequencing, only the plus strand of the transcriptome was queried with BLASTx. Transdecoder-predicted and translated ORFs were queried with BLASTp. The E-value threshold was set to 0.001, and only the top-scoring hit was retained. HMMER v. 3.1b2 (Finn et al., 2011) was used to search for conserved protein domains in predicted ORFs against the pfam-A (Finn et al., 2014) database (current as of Jan, 2015). BLAST homologies and Pfam domain entries were loaded into the pre-formatted Trinotate SQLite database, which contained UniProt-associated annotation information.

The complete, Trinity-assembled transcriptome (*Raw* assembly) was filtered to retain only transcripts expressed at an abundance of 1 FPKM (fragments per kilobase per transcript per million mapped reads; *Filtered* assembly) or only BLASTx-annotated transcripts (*BLASTx* assembly). Assembly statistics were calculated, and transcript coverage of each top-scoring unique hit was determined with custom scripts that came bundled with Trinity software. A local installation of the Galaxy platform (Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010) was used for manipulation of large datasets.

**Transcript abundance and differential expression analysis**

Transcript-level rather than unigene-level expression was investigated in order to retain isoform-specific information. Reads from individual libraries were aligned to the reference transcriptome with bowtie v. 1.1.1. (Langmead et al., 2009) and quantified by RSEM v. 1.2.18 (Li and Dewey, 2011). A table of TMM (Trimmed Mean of M-values)-normalized FPKM expression values and a separate table of raw fragment counts were generated with custom scripts.

Differentially expressed transcripts were identified from raw counts with the Bioconductor package EdgeR v. 3.1 (Robinson et al., 2010) in the statistical program R (R Development Core Team, 2008). Three biological replicates for each condition were provided. A subset of differentially expressed transcripts (FDR $< 0.001$ and fold change $\geq 4$) was extracted and used to generate a heatmap of hierarchically-clustered, $\log_2$-transformed and median-centered FPKM values. All scripts came bundled with Trinity software and default parameters were used, supplemented with the strand-specific parameter --SS_lib_type RF when applicable.

**Gene ontology analysis**

Blast2GO v. 3.0 (Conesa et al., 2005) was used to map GO terms to parent plant GOSlim terms in order to obtain a broad overview of the transcriptome. To identify enriched terms, a Fisher's test was conducted in Blast2GO with FDR $< 0.001$. For NAT GO term enrichment, any NAT pair with at least one protein-coding transcript, as annotated by BLASTx, was included in the test set. For enrichment of JA-responsive NAT pairs, in addition to the above protein-coding requirement, only pairs with differentially expressed sense and antisense transcripts were included in the test set. The raw, Trinity-assembled pokeweed transcriptome served as the reference set for all Fisher's tests conducted in this study.

**Identification of PAP isoforms and natural antisense transcripts**

Following transcriptome annotation, any transcript that matched a published PAP sequence as its top BLASTx hit and contained a predicted RIP protein domain was considered to be a PAP isoform. To identify NATs, a BLASTn search was conducted whereby the plus strand of the complete transcriptome was aligned to a local database containing the reverse complement of all transcripts. The BLASTn E-value threshold was set to 0.001.

**Cloning of novel PAP isoform and PAP natural antisense transcript**

All primer sequences used in this study are available in **Data Sheet 1**. Reverse transcription was performed on 500 ng of total pokeweed RNA in a 20 µL reaction volume containing 5 mM DTT, 1 µM reverse primer, 1X First Strand Buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 0.5 mM dNTPs, 20 units Murine RNase Inhibitor (NEB), and 25 units Superscript III reverse transcriptase (Thermofisher). The reaction was incubated at 42°C for 1 hour and heat inactivated at 70°C for 20 minutes.

Following cDNA synthesis, a PCR reaction was conducted, containing 1X Q5 buffer (NEB), 0.5 µM forward primer, 0.5 µM reverse primer, 200 mM dNTPs, 2 µL cDNA, and 1 unit Q5 DNA polymerase (NEB), in a total volume of 50 µL. The PCR program included an initial denaturation of 95°C for 30s, 30 cycles of 95°C for 30s, 58°C for 30s, 72°C for 120s, and finished with an extension at 72°C for 180s. PCR products were separated on low-melt agarose, and the correct size band excised and purified with EZ-10 Spin columns (Biobasic). The purified product was digested with BamHI and SalI, ligated into pBS-KSII, and sequenced.

**qRT-PCR validations**

For qRT-PCR, the reverse transcription step was performed in the same manner as cloning except that 2 µg of total pokeweed RNA from either control or JA treated plants was used with reverse primers corresponding to a specific transcript or 28S rRNA as the internal control. The qPCR reaction contained 5 µL of cDNA, 0.7 µM forward primer, 0.7 µM reverse primer, and 1X SYBR Green Mastermix (Clontech). Each reaction was split into three technical replicates and analyzed in a Qiagen Rotor-gene-Q real time PCR cycler. Ct values were calculated with the ΔΔCt relative quantification method. Three biological replicates were conducted for each transcript. For statistical analysis, a one-tailed, unpaired Student's t-test was conducted using GraphPad Prism v.5.01.

## 2.4 Results

**Assembly and annotation of the pokeweed mRNA transcriptome**

An overview of the entire study is provided in **Supplementary Figure 1**. A total of 406,995,054 high-quality reads from control and JA-treated pokeweed plants were combined, and the mRNA transcriptome was assembled with Trinity software. Assembly statistics are provided in **Table 1**. The complete pokeweed transcriptome (*Raw*) contained 216,891 transcripts belonging to 177,709 unigenes. To identify contigs expressed at a reasonable threshold, the *Raw* assembly was filtered to retain only those having a minimum abundance of 1 FPKM; this reduced the number of transcripts and unigenes to 89,682 and 77,731, respectively (*Filtered*). The *Raw* assembly was also filtered on the basis of BLASTx annotation, which resulted in 59,096 and 38,291 annotated transcripts and unigenes, respectively (*BLASTx*). Furthermore, 16,245 unique proteins were represented by transcripts with at least 70% BLASTx alignment coverage, indicating a high-

**Table 1. Assembly statistics for pokeweed mRNA transcriptomes.**

| Parameter | Raw | Filtered | BLASTx |
|---|---|---|---|
| # of Transcripts | 216,891 | 89,682 | 59,096 |
| # of Unigenes | 177,709 | 77,731 | 38,291 |
| Transcriptome size (Mb) | 157 | 74 | 85 |
| N50 (bp) | 1,168 | 1,617 | 2,102 |
| Average length (bp) | 724 | 821 | 1,439 |
| Median length (bp) | 406 | 373 | 1,159 |
| Minimum length (bp) | 201 | 201 | 201 |
| Maximum length (bp) | 15,776 | 15,776 | 15,776 |
| # of Transcripts ≥ 1Kb | 42,541 | 24,436 | 32,746 |

*Raw: the complete, Trinity-assembled transcriptome; Filtered: only transcripts ≥ 1FPKM were retained; BLASTx: only BLASTx-annotated transcripts from the UniRef90 and SwissProt databases were retained.*

quality transcriptome assembly (**Data Sheet 2**). The Trinotate report for BLASTx-annotated transcripts is available in **Data Sheet 3**. Interestingly, only 38% of transcripts in the *Filtered* assembly were also BLASTx annotated, suggesting that the majority of pokeweed-expressed transcripts are not shared with those from available plants.

To assess the contiguity of our different assemblies, we also determined their N50 values (**Table 1**). The N50 statistic is a weighted median, such that 50% of the assembly is contained in contigs equal to or larger than this value. N50 values for the *Raw, Filtered*, and *BLASTx* assemblies were 1,168, 1,617, and 2,102 bp, respectively. These values indicated that expressed and annotated transcripts tended to be longer in length, which was confirmed from the assembly length distributions (**Figure 6**). Notably, the *Raw* assembly had a majority of transcripts between 200-600 bp in length, while this size class accounted for a smaller proportion of the *Filtered* and *BLASTx* assemblies. These short transcripts could represent partial transcripts, microRNA precursors, and/or assembly artefacts.

Functional annotation of the complete, assembled pokeweed transcriptome was carried out with the Trinotate pipeline, and GO analysis was conducted using Blast2GO software. GO terms were mapped to corresponding plant GOSlim terms in order to obtain a summary of the transcriptome (**Supplementary Figure 2**). GOSlims are cut-down versions of the ontology with reduced detail of lower-level terms; they are useful for providing a general overview of the transcriptome. In total, 36,423 transcripts were annotated with 238,251 GO terms, distributed amongst the categories of Biological Process, Molecular Function, and Cellular Component. Within the Biological Process category, the most abundant terms included response to stress (5,471 transcripts; 15.0% of all transcripts), transport (5,365; 14.7%), and cellular protein modification process (5,200; 14.3%). Nucleotide binding (8,403; 23.1%), DNA binding (4,415; 12.1%), and

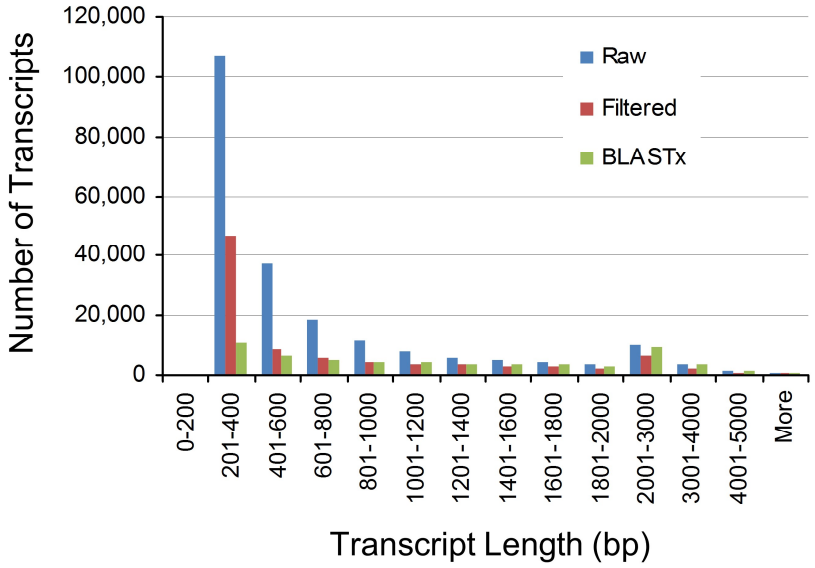



**Figure 6. Length distributions of pokeweed mRNA transcriptomes.**

kinase activity (3,219; 8.8%) comprised the majority of terms in the Molecular Function category. Finally, the Cellular Component distribution indicated that most annotated proteins localized to the plasma membrane (4,974; 13.7%), plastid (3,559; 9.8%), or cytosol (2,887; 7.9%). Taken together, GO analysis indicates that the majority of pokeweed transcripts can be grouped into a small number of broad yet distinct functional categories.

**Identification of jasmonic acid-responsive genes**

Following transcriptome annotation, our next goal was to identify JA-responsive genes through differential expression analysis. Briefly, reads from each library were individually aligned back to the complete reference transcriptome, and the abundance of each transcript was determined. Differential expression analysis was then conducted based on normalized read counts. The most abundant, BLASTx-annotated transcripts from control and JA-treated plants are summarized in **Table 2,** and abundances of all transcripts are available in **Data Sheet 4**. Control plants had high expression of several genes encoding photosynthetic proteins, including RuBisCO and photosystem-associated components. Conversely, JA-treated plants had high abundance of transcripts encoding defense proteins, including two defensin-like proteins, two isoforms of PAP, and a proteinase, in addition to constitutive plant metabolic proteins.

A total of 8,264 transcripts were differentially expressed between control and JA-treated plants (FDR < 0.05; **Supplementary Figure 3**). A subset of 2,770 (FDR < 0.001, fold change ≥ 4) was defined for downstream analysis (**Figure 7A**). Of these, the majority of transcripts increased with JA treatment (2067; 75%). Furthermore, as shown in **Figure 7B**, most of the transcripts were expressed in both control and JA-treated plants (2192; 79%); interestingly, a considerable number was detected only in JA-treated plants (434; 16%). Of these JA treatment-

**Table 2. Most abundant, BLASTx-annotated transcripts expressed in pokeweed under control (E) and JA treatments.**

| Transcript ID | Top BLASTx Hit | Gene Name | $log_2$FPKM |
|---|---|---|---|
| **E** | | | |
| c58232_g7_i1 | RBS1_MESCR | Ribulose bisphosphate carboxylase small chain 1, chloroplastic | 14.74 |
| c58494_g1_i1 | CB2A_SPIOL | Chlorophyll a-b binding protein, chloroplastic | 14.73 |
| c111935_g1_i1 | PSBR_SOLTU | Photosystem II 10 kDa polypeptide, chloroplastic | 14.00 |
| c58232_g6_i1 | RBS2_MESCR | Ribulose bisphosphate carboxylase small chain 2, chloroplastic | 13.83 |
| c111786_g1_i1 | GL33_ARATH | Germin-like protein subfamily 3 member 3 | 13.09 |
| c47181_g1_i1 | GRP1_DAUCA | Glycine-rich RNA-binding protein | 12.64 |
| c16825_g1_i2 | CAHC_SPIOL | Carbonic anhydrase, chloroplastic | 12.58 |
| c111825_g1_i1 | CB23_TOBAC | Chlorophyll a-b binding protein 36, chloroplastic | 12.57 |
| c60753_g1_i1 | CB12_PETHY | Chlorophyll a-b binding protein, chloroplastic | 12.27 |
| c111752_g1_i1 | PSAK_ARATH | Photosystem I reaction center subunit psaK, chloroplastic | 12.18 |
| **JA** | | | |
| c112185_g1_i1 | DEF_NELNU | Defensin-like protein | 16.54 |
| c51788_g1_i1 | DF322_SOLTU | Defensin-like protein P322 | 15.30 |
| c58494_g1_i1 | CB2A_SPIOL | Chlorophyll a-b binding protein, chloroplastic | 14.68 |
| c58232_g7_i1 | RBS1_MESCR | Ribulose bisphosphate carboxylase small chain 1, chloroplastic | 14.37 |
| c3137_g1_i1 | RIP1_PHYAM | Antiviral protein I | 14.35 |
| c111935_g1_i1 | PSBR_SOLTU | Photosystem II 10 kDa polypeptide, chloroplastic | 14.01 |
| c58232_g6_i1 | RBS2_MESCR | Ribulose bisphosphate carboxylase small chain 2, chloroplastic | 13.80 |
| c60978_g1_i1 | XCP1_ARATH | Xylem cysteine proteinase 1 | 13.49 |
| c3192_g1_i1 | RIP2_PHYAM | Antiviral protein 2 | 13.18 |
| c47181_g1_i1 | GRP1_DAUCA | Glycine-rich RNA-binding protein | 12.70 |

*Transcripts are listed in order of decreasing abundance.*

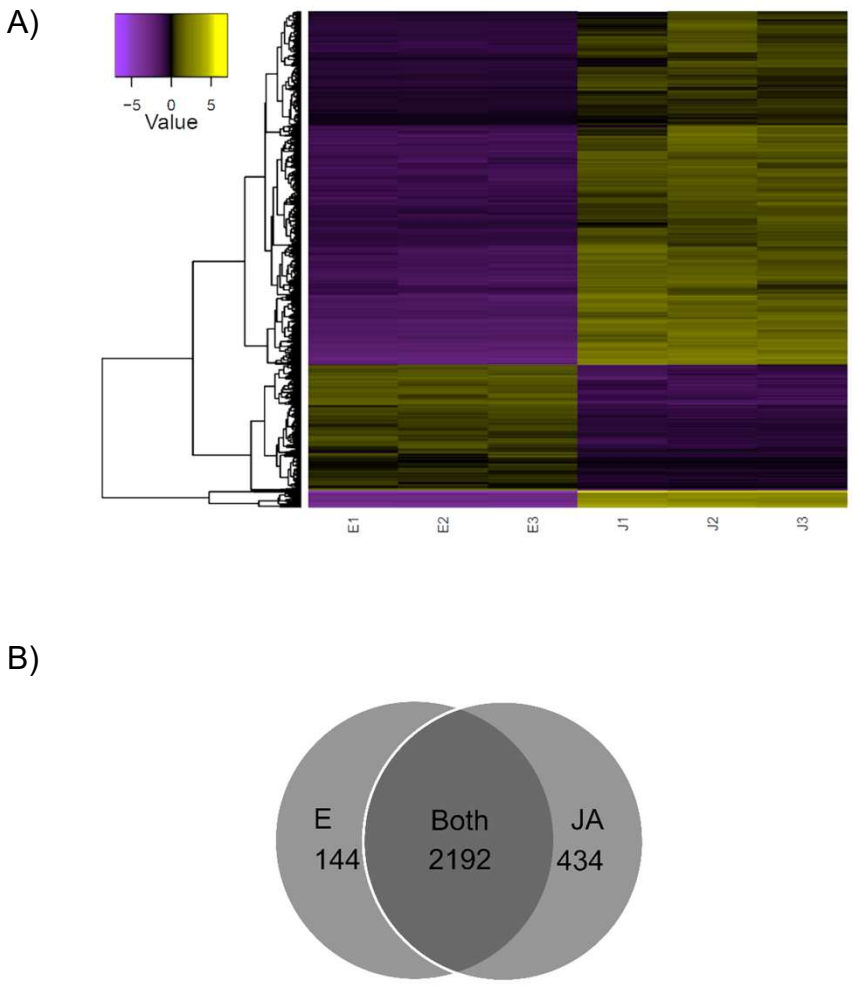



**Figure 7. Identification of JA-responsive genes in pokeweed.** **A)** Heat map of expression values ($log_2$FPKM, median-centered) of the top differentially expressed transcripts (FDR < 0.001, fold change ≥ 4). **B)** Venn diagram depicting treatment-specific expression patterns of transcripts in A).

specific transcripts, 165 encoded known proteins as annotated by BLASTx (**Data sheet 5**). Many have well-established roles in defense, including a pathogenesis-related protein and several chitinases, proteinases, peroxidases, and terpenoid biosynthesis enzymes. JA-induced transcription factor families such as ERF, MYB, and TIFY were identified, as well as two enzymes involved in JA biosynthesis: jasmonate O-methyltransferase and 4-coumarate-CoA ligase-like 5. Interestingly, among the most abundant, JA-specific transcripts was a putatively novel isoform of PAP, c115037_g1_i1.

The top JA-responsive, BLASTx-annotated transcripts are summarized in **Table 3**, and the complete list of differential expression results is provided in **Data Sheet 6**. As expected, many of these transcripts encoded proteins involved in JA metabolism (**Supplementary Figure 4**). Additionally, several defense genes were among those most differentially expressed; these included intracellular ribonuclease LX, nerolidol synthases, antiviral protein alpha, and defensin-like protein. To obtain insight into the functional roles of JA-responsive transcripts in pokeweed, we conducted GO term enrichment analysis (**Table 4**). Up- and down-regulated transcripts were investigated separately in order to determine their independent contributions within the plant. Up-regulated transcripts were highly enriched in terms related to stress and defense responses, indicating marked transcriptional reprogramming in JA-treated plants. Down-regulated transcripts were not enriched in any well-defined stress responses, and the enriched terms did not appear to represent any common themes.

To validate RNA-seq results, the expression of eight randomly selected transcripts and two selected transcripts (discussed below) from our defined subset was assessed by qRT-PCR from control and JA-treated plants (**Figure 8**). An $R^2$ correlation value of 0.917 was obtained, indicating high correspondence between the two methods of transcript quantitation.

**Table 3. Top JA-responsive, BLASTx-annotated genes in pokeweed.**

| Transcript ID | Top BLASTx Hit | Gene name | $log_2$ FC | $log_2$ FPKM E | $log_2$ FPKM JA | FDR |
|---|---|---|---|---|---|---|
| c112223_g1_i1 | LOX21_SOLTU | Linoleate 13S-lipoxygenase 2-1, chloroplastic | 13.97 | -4.63 | 9.58 | 1.51E-182 |
| c53706_g1_i2 | TRPB_CAMAC | Tryptophan synthase beta chain 2, chloroplastic | 13.36 | -4.18 | 9.32 | 5.28E-157 |
| c20579_g1_i1 | RNLX_SOLLC | Intracellular ribonuclease LX | 13.84 | -2.93 | 11.08 | 3.64E-148 |
| c112209_g1_i1 | NATT3_THANI | Natterin-3 | 8.67 | 1.72 | 10.51 | 8.82E-144 |
| c3273_g2_i1 | NES2_FRAAN | (3S,6E)-nerolidol synthase 2, chloroplastic/mitochondrial | 10.63 | -2.41 | 8.34 | 8.92E-137 |
| c61047_g1_i1 | BSPA_POPDE | Bark storage protein A | 13.10 | -4.08 | 9.16 | 2.85E-135 |
| c3273_g1_i1 | NES1_FRAAN | (3S,6E)-nerolidol synthase 1 | 13.30 | -6.42 | 7.29 | 6.16E-135 |
| c50513_g1_i1 | ZOG_PHALU | Zeatin O-glucosyltransferase | 13.77 | -7.38 | 7.00 | 3.57E-125 |
| c60944_g1_i1 | RIPA_PHYAM | Antiviral protein alpha | 12.99 | -1.58 | 11.68 | 9.32E-120 |
| c112185_g1_i1 | DEF_NELNU | Defensin-like protein | 13.19 | 3.30 | 16.54 | 3.77E-119 |

*Transcripts are listed in order of decreasing significance.*

**Table 4. Top enriched GO terms for JA-responsive genes in pokeweed (FDR < 0.001).**

| Down-regulated transcripts (703) | | Up-regulated transcripts (2067) | |
|---|---|---|---|
| **GO Term** | **FDR** | **GO Term** | **FDR** |
| extracellular region | 5.96E-04 | response to wounding | 1.28E-26 |
| hydrolase activity, hydrolyzing O-glycosyl compounds | 2.88E-03 | response to endogenous stimulus | 1.82E-16 |
| heme binding | 2.88E-03 | response to jasmonic acid | 5.82E-16 |
| methylammonium transmembrane transporter activity | 2.88E-03 | response to chitin | 3.90E-15 |
| cell wall | 3.91E-03 | defense response | 1.54E-14 |
| storage vacuole | 3.91E-03 | regulation of systemic acquired resistance | 2.60E-10 |
| nitrate reductase (NADH) activity | 1.31E-02 | cytoplasm | 2.60E-10 |
| nucleus | 1.73E-02 | cellular component organization or biogenesis | 6.27E-10 |
| molybdopterin cofactor binding | 1.98E-02 | heme binding | 7.75E-10 |
| oxidation-reduction process | 2.02E-02 | nutrient reservoir activity | 1.19E-09 |

*The number in brackets indicates the number of down- or up-regulated transcripts within the test set.*

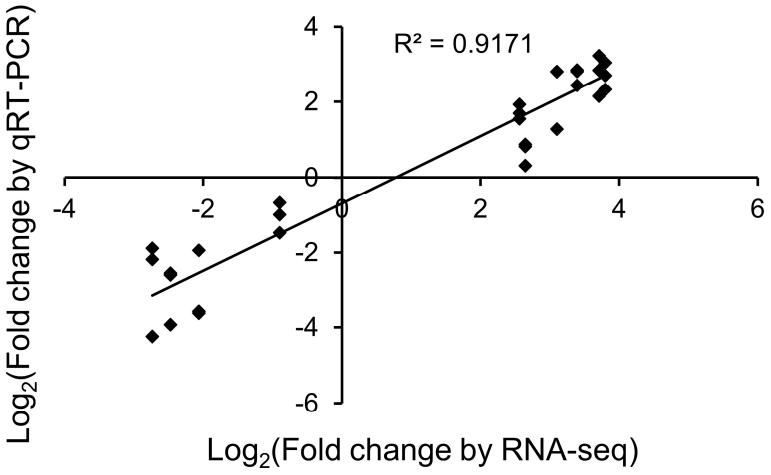



**Figure 8. Validation of RNA-seq differential expression results.** The correlation of JA-induced expression changes obtained from RNA-seq and qRT-PCR is shown for ten transcripts, eight of which were randomly selected. Results for qRT-PCR are from three independent biological replicates for each transcript.

**Identification and analysis of PAP isoforms**

Following transcriptome annotation, we were able to identify assembled PAP isoforms on the basis of homology with published sequences. Six transcripts were annotated as RIPs through BLASTx and contained predicted RIP domains within their translated open reading frames (ORFs); their annotation information is summarized in **Table 5**. Two isoforms, PAP-I and PAP-II, had perfectly assembled ORFs. Partial transcripts of PAP-alpha and PAP-S were also identified, with 73% and 39% coverage and nearly 100% sequence identity with their respective hits. Interestingly, transcript c18776_g1_i1 had only 40% identity and 86% coverage with PAP-alpha, its top BLASTx hit. Furthermore, the E-value of its identified RIP domain was more significant than that of PAP-II, which was correctly assembled. We cloned and sequenced the predicted ORF of c18776_g1_i1 from pokeweed total RNA, confirming expression of the transcript in the plant. Taken together, we hypothesize that transcript c18776_g1_i1 represents a novel PAP isoform. Transcript c115037_g1_i1, with 81% identity to PAP-I, may be another novel isoform; however, with only 38% coverage, we could not rule out the possibility that this transcript was an assembly artefact.

Differential expression results of the six identified PAP isoforms are provided in **Table 6**. With the exception of c18776_g1_i1, the hypothesized novel isoform, all other transcripts showed a significant increase in abundance upon JA treatment (FDR < 0.05). PAP-S had the most significant result, with a $\log_2$ fold change of 12.07, a remarkable 4,300-fold increase. Transcript c115037_g1_i1, the other putative novel isoform, showed the highest $\log_2$ fold change of 13.17. The expression patterns of c18776_g1_i1 and c115037_g1_i1 are distinct from PAP-alpha and PAP-I, their respective top hits, further supporting the hypothesis that they are novel isoforms. We validated the expression of c18776_g1_i1 by qRT-PCR. Although its reduction with JA treatment

**Table 5. Identification of PAP isoforms in the pokeweed mRNA transcriptome from BLASTx and Pfam annotations.**

| Transcript | | BLASTx Alignment | | | | | | RIP Protein Domain (PF00161.14) | |
|---|---|---|---|---|---|---|---|---|---|
| **ID** | **Length (bp)** | **Top Hit** | **Gene Name** | **Region in Query (Q) and Hit (H)** | **E Value** | **% Identity** | **% Coverage** | **Region** | **E Value** |
| c3137_g1_i1 | 1259 | RIP1_PHYAM | Antiviral protein I | Q:129-1067, H:1-313 | 0 | 100.00 | 100.00 | 28-239 | 3.50E-67 |
| c3192_g1_i1 | 1401 | RIP2_PHYAM | Antiviral protein 2 | Q:173-1102, H:1-310 | 0 | 98.71 | 100.00 | 29-240 | 2.70E-58 |
| **c18776_g1_i1** | **1479** | **RIPA_PHYAM** | **Antiviral protein alpha** | **Q:285-1019, H:26-277** | **2E-39** | **39.84** | **85.71** | **39-242** | **2.50E-59** |
| c30332_g1_i2 | 643 | RIPA_PHYAM | Antiviral protein alpha | Q:2-643, H:62-275 | 5E-156 | 99.53 | 72.79 | 1-180 | 1.90E-53 |
| c16017_g1_i1 | 305 | RIPS_PHYAM | Antiviral protein S | Q:2-304, H:85-185 | 8E-67 | 99.01 | 38.70 | 1-101 | 6.10E-24 |
| **c115037_g1_i1** | **552** | **RIP1_PHYAM** | **Antiviral protein I** | **Q:2-358, H:49-167** | **9E-79** | **80.67** | **38.02** | **2-106** | **3.90E-20** |

*Putative novel PAP isoforms are indicated in bold.*

**Table 6. Differential expression of PAP isoforms.**

| Transcript ID | Top BLASTx Hit | $log_2$FC | $log_2$FPKM E | $log_2$FPKM JA | FDR |
|---|---|---|---|---|---|
| c16017_g1_i1 | RIPS_PHYAM | 12.07 | -1.79 | 10.55 | 1.87E-96 |
| **c115037_g1_i1** | **RIP1_PHYAM** | **13.17** | **-3.32** | **7.16** | **1.37E-75** |
| c30332_g1_i2 | RIPA_PHYAM | 7.77 | -2.20 | 5.62 | 2.12E-42 |
| c3137_g1_i1 | RIP1_PHYAM | 3.83 | 10.41 | 14.35 | 1.76E-39 |
| c3192_g1_i1 | RIP2_PHYAM | 3.11 | 9.97 | 13.18 | 3.95E-26 |
| **c18776_g1_i1** | **RIPA_PHYAM** | **-0.90** | **7.69** | **6.88** | **6.81E-02** |

*Transcripts are listed in order of decreasing significance. Putative novel isoforms are indicated in bold.*

was not significant by RNA-seq (FDR < 0.05), qRT-PCR indicated a $\log_2$ fold change of -1.04 (0.49 fold), which was significant (Student's t test, $p < 0.05$; **Figure 8**). The fully-assembled PAP-I and PAP-II isoforms showed the lowest $\log_2$ fold changes of 3.83 and 3.11, respectively; interestingly, they were also the most abundant. Taken together, these results indicate that PAP isoforms respond differently to JA treatment, in terms of fold change and abundance.

**Discovery of natural antisense transcripts in pokeweed**

Owing to strand-specific sequencing of the pokeweed mRNA transcriptome, we were able to identify putative NATs. This involved performing a BLASTn search of the assembled transcriptome against its reverse complement to find transcript pairs having significant sequence complementarity (E < 0.001). In total, 52,305 NAT pairs were detected, although this number is an under-representation because only the best-matching partner for each transcript was retained in our analysis.

We conducted GO enrichment analysis to investigate the potential functional roles of NATs in pokeweed. As shown in **Figure 9**, NATs were enriched in 24 GO terms relative to all pokeweed transcripts from the raw assembly (FDR < 0.001). Of these, "chloroplast stroma" was most significant, and three other chloroplast-related terms were also enriched: "chloroplast thylakoid membrane", "chlorophyll biosynthetic process", and "chloroplast thylakoid lumen". From the 52,305 pairs of NATs, 2,502 transcripts were differentially expressed (FDR < 0.001, fold change ≥ 4). Of the differentially expressed NATs, 88 pairs involved both partners, suggesting co-regulation in response to JA. The 88 pairs were enriched in three GO terms relative to all pokeweed transcripts: "nitrate assimilation", "nitrate reductase (NADH) activity", and "molybdopterin cofactor binding" (FDR < 0.001). The transcripts relating to these terms were annotated as

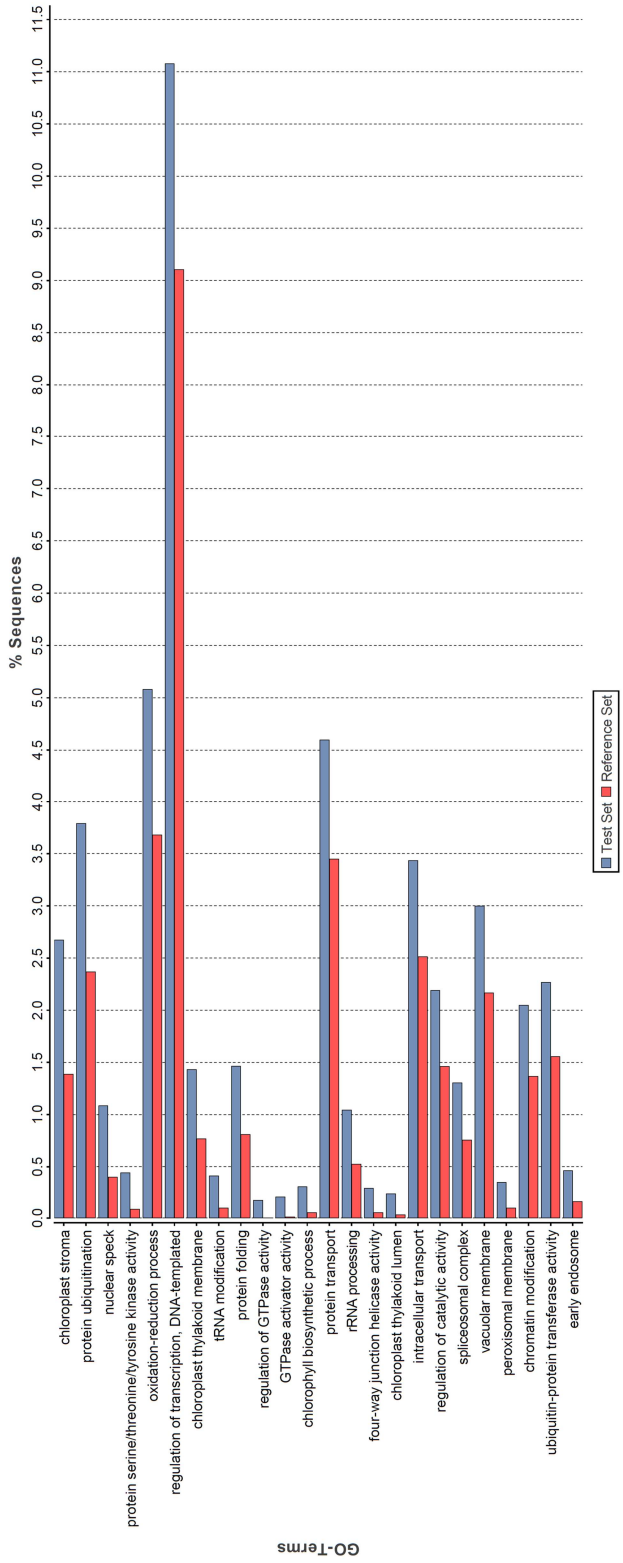



**Figure 9. Functional analysis of natural antisense transcripts in pokeweed.** GO enrichment analysis was conducted on all NATs (Test Set) against the raw pokeweed transcriptome assembly (Reference Set). Enriched terms (FDR < 0.001) are listed in order of increasing significance (bottom to top).

NRT1/ PTR FAMILY members and nitrate reductases. Therefore, it appears that in pokeweed, NATs in general are important in the chloroplast, while JA-responsive NATs may regulate nitrate metabolism.

As shown in **Table 7**, four PAP isoforms had significant NAT counterparts (PAP-I, PAP-II, and the two potentially novel isoforms). Each NAT showed perfect alignment identity and nearly full-length coverage with its respective PAP isoform. Interestingly, the expression of PAP isoforms and their corresponding NATs showed a strong positive correlation, with an $R^2$ value of 0.909 (**Figure 10**). We cloned and sequenced the PAP-I NAT (c61645_g1_i1) from pokeweed total RNA to validate its expression. Furthermore, qRT-PCR of PAP-I NAT from JA-treated plants indicated a significant increase of 6.68 fold ($log_2$FC of 2.74; Student's t test, $p < 0.05$; **Figure 8**). These results confirm the expression of a JA-responsive transcript that is antisense to the PAP-I sequence. In addition, they suggest that certain PAP isoforms may be regulated by NATs, which would constitute a novel form of PAP gene regulation.

## 2.5 Discussion

Through transcriptome assembly, annotation, and differential expression analysis, we identified genes that are significantly affected by JA and could mediate defense against pathogens and herbivores in *P. americana*. We validated the existence of several previously reported PAP isoforms and characterized their differential expression patterns. This work also led to the discovery of a potentially novel PAP isoform and identification of NATs that may regulate PAP gene expression.

**Table 7. Identification and expression of putative PAP natural antisense transcripts.**

| PAP Isoform Transcript | NAT | NAT Length (bp) | % ID | % Coverage | Aligned regions of PAP (P) and NAT (N) | $log_2$ FC | $log_2$ FPKM E | $log_2$ FPKM JA | FDR |
|---|---|---|---|---|---|---|---|---|---|
| c3137_g1_i1 | c61645_g1_i1 | 1260 | 100 | 100.08 | P:22-1259, N:1-1238 | 3.71 | 5.08 | 8.90 | 1.95E-27 |
| c3192_g1_i1 | c37772_g1_i1 | 1180 | 99.92 | 84.23 | P:85-1264, N:1-1180 | 3.05 | 3.64 | 6.80 | 1.16E-17 |
| c18776_g1_i1 | c24011_g1_i1 | 1319 | 100 | 89.18 | P:43-1361, N:1-1319 | -0.92 | 1.76 | 0.93 | 0.35936 |
| c115037_g1_i1 | c21462_g1_i2 | 551 | 100 | 99.82 | P:2-552, N:1-551 | 7.18 | -3.32 | 1.15 | 2.44E-06 |

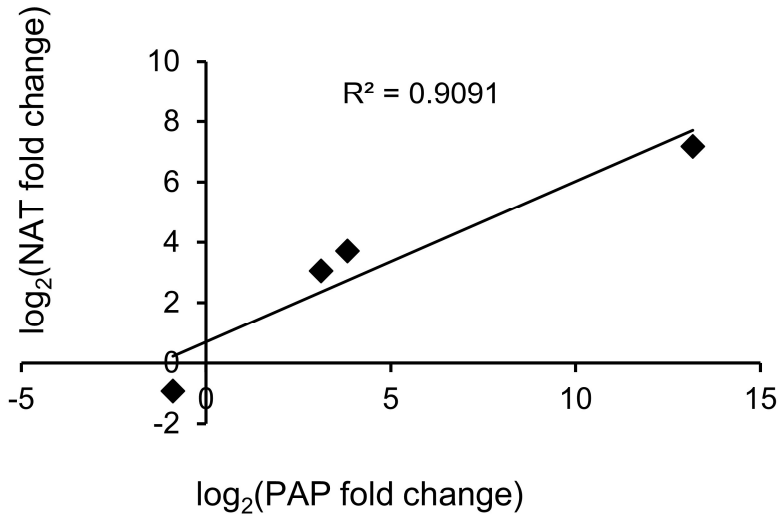



**Figure 10. Expression of PAP sense and antisense transcripts.** The correlation of JA-induced expression changes obtained by RNA-seq is shown for each PAP isoform and its corresponding natural antisense transcript.

**Construction of a high-quality pokeweed reference transcriptome**

We generated a robust pokeweed mRNA transcriptome by combining strand-specific, paired-end sequencing reads from biological replicates of JA-treated and control plants. Over 400 million processed reads were leveraged for assembly, which is considered very deep sequencing (Haas et al., 2013). Comparison of our assembly with sugar beet, a well-studied plant from the same taxonomic order as pokeweed, reveals important similarities to help validate our results. For example, *de novo* assembly of the sugar beet mRNA transcriptome yielded a total of 225,385 transcripts from 165,742 unique loci, providing an N50 value of 1,185 bp (Mutasa-Göttgens et al., 2012). These statistics are comparable to our pokeweed *Raw* assembly, which had 216,891 transcripts, 177,709 unigenes, and an N50 value of 1,168 bp. Furthermore, others reported that approximately 80% of sugar beet unigenes are between 200 and 500 bp in length (Fugate et al., 2014), in agreement with the abundance of short transcripts we detected in pokeweed. The enrichment of short transcripts seems to be a general trend amongst *de novo* assembled transcriptomes, perhaps due to the assembly algorithm which relies on short overlapping sub-sequences known as k-mers (Ranjan et al., 2014). Through BLASTx searches against public sequence databases, we were able to annotate 59,096 transcripts. This is in line with other studies of non-model plants, with examples of approximately 40-50,000 annotated transcripts or unigenes (Farrell et al., 2014; Jin et al., 2014; Ranjan et al., 2014). Taken together, reports from other plants support the validity of our pokeweed assembly.

**Discovery of putative defense genes in pokeweed**

We were interested in identifying JA-responsive genes in pokeweed because this hormone mediates broad-spectrum defense. We reported 165 annotated transcripts specific to JA treatment.

Many of these transcripts encoded factors involved in JA biosynthesis, JA signalling, and JA-mediated defense, which have been identified in other transcriptome-wide studies involving transgenic plants deficient in JA biosynthesis or signalling (Abuqamar et al., 2006; Ku et al., 2011; Yan et al., 2014). Identification of these JA responses allows us to integrate PAP within broader defence pathways of pokeweed.

We also sought to identify genes other than PAP that may contribute to defense in this plant. One interesting candidate was intracellular ribonuclease LX, which showed a 14,664-fold increase with JA. Such changes are not uncommon in plants responding to biotic and abiotic stresses (Li et al., 2014; Liu et al., 2013; Pierce and Rey, 2013). This gene, characterized in tomato, is involved in programmed cell death responses including senescence and is thought to contribute to nutrient recycling by mediating RNA turnover (Lehmann et al., 2001; Lers et al., 2006). It is well-established that jasmonate is associated with localized cell death as part of the hypersensitive response in plants, providing a possible explanation for the increased expression of this gene in JA-treated plants. Future work will investigate if this enzyme has a role in pokeweed defense, as pathogen-responsive RNases have been reported in other plants (Ayashi et al., 2003; Galiana et al., 1997).

Another markedly up-regulated gene in our study was annotated as bark storage protein A and is involved in nitrogen storage in senescing leaves (Coleman et al., 1991; Wetzel et al., 1989). Storage proteins have previously been linked to jasmonate-mediated responses, such as VSP2 in *Arabidopsis*, an anti-insect acid phosphatase (Liu et al., 2005). A final candidate is defensin-like protein, which in addition to being differentially expressed, was also the most abundant transcript in JA-treated plants. In *Arabidopsis*, the defensin PDF1.2 is a well-established marker of the ethylene and jasmonate signalling pathway that mediates resistance to pathogens (Manners et al.,

1998; Penninckx et al., 1998). Identification of JA-responsive genes in pokeweed will allow a broader characterization of factors involved in defense in this plant. Furthermore, since only 38% of expressed pokeweed transcripts were annotated, this non-model plant transcriptome represents a source of new genes that could have agricultural benefits.

**Characterization of PAP isoforms and their differential expression patterns**

Through transcriptome annotation, we confirmed the presence of the following isoforms in pokeweed leaf tissue from 4-leaf plants: PAP-I, PAP-II, PAP-S and PAP-alpha. Based on previous reports, PAP-I is expressed in spring leaves (Irvin, 1975), PAP-II in early summer leaves (Irvin and Kelly, 1980), PAP-S in seeds (Barbieri et al., 1982), and PAP-alpha in various tissues (Kataoka et al., 1992). We also presented sequence and qRT-PCR evidence to support the existence of a novel PAP isoform. Our finding that PAP-I was the most abundant isoform in young plants, followed by PAP-II, agrees with their documented temporal profiles. Furthermore, the high abundance of the isoforms corresponds with the report that PAP comprises up to 0.5% of total soluble protein in leaves (Bonness et al., 1994). Interestingly, we only found one annotated PAP-S isoform. This disagrees with a previous finding that PAP-S is a mix of two seed isoforms, PAP-S1 and PAP-S2, having 85% nucleotide sequence identity (Honjo et al., 2002). Our result was not likely due to the assembly algorithm, since Trinity is able to differentiate between sequences that are up to 95% identical (Grabherr et al., 2011). This does not preclude the possibility that two PAP-S isoforms exist in seeds, as we only investigated leaf tissue. Furthermore, we did not identify PAP-R or PAP-III isoforms, which have been purified from root tissue and late summer leaves, respectively (Bolognesi et al., 1990; Rajamohan et al., 1999).

RNA-seq analysis allowed us to compare the expression patterns of multiple PAP isoforms simultaneously. This level of distinction in a single experiment has not yet been possible, likely due to difficulties related to high sequence identity amongst isoforms. With the exception of PAP-II, the nucleotide sequences of isoforms are >70% identical. In the current study, we found that all isoforms other than c18776_g1_i1 were significantly up-regulated with JA. PAP isoforms exhibited differences in terms of abundance and fold change. Such differences could indicate that they have specialized roles in defense, or more generally within the plant. For example, analysis of two RIP isoforms in spinach revealed that one was highly expressed in embryos and not responsive to salicylic acid, while the other showed weak expression throughout the plant but was induced with salicylic acid (Kawade and Masuda, 2009). Authors suggested that the two RIPs may have different functions, one in embryogenesis and the other in defense. In pokeweed, PAP isoforms differ in their ability to depurinate eukaryotic and prokaryotic ribosomes *in vitro* (Honjo et al., 2002). A comprehensive RNA-seq study that includes different developmental stages and stresses would allow further insight into the possible roles of PAP isoforms.

**Natural antisense transcripts in pokeweed**

NATs are pairs of endogenous transcripts with high sequence complementarity, capable of forming double-stranded RNA and affecting gene expression in *cis* or *trans*. *Cis*-NATs are transcribed from the same genomic locus in opposite orientations and have perfect sequence complementarity, while *trans*-NATs are derived from different loci. Our sequence analysis identified over 52,000 NAT pairs in pokeweed. A recent study identified 37,238 NAT pairs in *Arabidopsis*, and an astonishing 70% of annotated mRNAs were associated with antisense transcripts (Wang et al., 2014). Furthermore, 60% of NAT pairs were comprised of fully

overlapping transcripts. In pokeweed, we found perfectly identical, fully overlapping antisense transcripts of PAP-I, PAP-II, c18776_g1_i1 and c115037_g1_i1. The fact that each NAT was perfectly complementary to its respective isoform suggests that they act in *cis*. We also found that PAP NATs and their corresponding sense transcripts had positively correlated responses to JA, indicating that NATs may regulate PAP expression. This agrees with studies in humans and other mammals, which report significant positive correlations between sense and antisense transcription (Ling et al., 2013; Oeder et al., 2007). Fewer genome-wide studies of NAT transcription have been performed for plants. In *Arabidopsis*, a study of light-responsive long non-coding NATs identified 626 positively correlated and 766 negatively correlated pairs (Wang et al., 2014). NATs have been proposed to modulate gene expression at various levels, including transcription, post-transcriptional small RNA interference, mRNA splicing, and RNA stability (Liu et al., 2015; Pelechano and Steinmetz, 2013; Zhang et al., 2013b). Future work will investigate mechanisms by which NATs may regulate PAP expression.

Apart from PAP regulation, we conducted preliminary investigations into the broader roles of NATs in pokeweed. As a group, NATs were enriched in several GO terms relating to the chloroplast. Although the implications of this are unknown, it agrees with a previous report from *Arabidopsis* that the chloroplast encodes a diverse group of antisense RNAs mapping to protein-coding genes (Hotto et al., 2011). Approximately 2,500 NATs were JA-responsive, and 88 pairs were considered biologically relevant because both sense and antisense strands were differentially expressed. These NAT pairs were enriched in GO terms relating to nitrate metabolism, and the involved transcripts were annotated as NRT1/ PTR FAMILY members and nitrate reductases. These enzymes have well-established roles in nitrate assimilation, which involves the reduction of nitrate to nitrite to ammonium, which is ultimately incorporated into amino acids for plant growth

(Dechorgnat et al., 2010). Importantly, the reduction of nitrite to ammonium occurs in chloroplasts, further supporting the connection between NATs and this specific organelle. Furthermore, NRT1/ PTR FAMILY members capable of transporting jasmonoyl-isoleucine, the bioactive form of jasmonate, were recently identified (Chiba and Shimizu, 2015). Given that plant stress decouples nitrate assimilation and photosynthesis, mediated by the jasmonate/ethylene signalling pathway (Zhang et al., 2014), NATs in pokeweed may be involved in regulating the trade-off between stress response and plant growth.

**Summary and relevance**

Here, we have assembled and annotated the pokeweed leaf mRNA transcriptome under JA treatment. In addition to the identification of many differentially expressed transcripts, we also characterized the expression of multiple PAP isoforms through RNA-seq analysis, including a potentially novel isoform of PAP. Finally, we presented the first report of NATs in the pokeweed plant, and confirmed expression of a PAP-NAT, which may indicate a novel form of RIP gene regulation. The pokeweed transcriptome will enable further investigations into the robust defense strategies of this species and other Phytolaccaceae family members. Heterologous expression of PAP has already been applied successfully to produce virus resistant plants (Dai et al., 2003; Lodge et al., 1993; Wang et al., 1998; Zoubenko et al., 1997). Pokeweed accumulates high levels of metal from contaminated soils, but little is known about the genes mediating this tolerance (Dou et al., 2009; Zhao et al., 2011, 2012). With a complete repository of pokeweed mRNA sequences, we anticipate the discovery of beneficial genes in this plant that could improve the resiliency of agricultural crops.

## 2.6 Acknowledgement

The authors gratefully acknowledge permission to reproduce the KEGG pathway image #00592 from *http://www.kegg.jp/kegg/kegg1.html* used in Supplementary Figure 4.

# CHAPTER 3:

## Integration of the pokeweed miRNA and mRNA transcriptomes reveals widespread targeting of jasmonic acid-responsive genes

Published manuscript

**Neller, K. C. M.**, Klenov, A., Guzman, J. C., and Hudak, K. A. (2018). *Front. Plant Sci.* 9, 589. doi: 10.3389/fpls.2018.00589.

**Contributions:** KN and KH conceived the study. KN performed bioinformatic analyses. AK conducted plant treatments and wet-lab validations (stem-loop qRT-PCR and 5′ RACE). JG determined miRNA/target expression correlations and generated the interaction network. KN drafted the manuscript. KH edited the manuscript.

All Supplementary Material referenced in this manuscript (**Supplementary Figures S1-S5**, **Supplementary Tables S1-S3**) is available at the above doi.

## 3.1 Abstract

The American pokeweed plant, *Phytolacca americana*, displays broad-spectrum resistance to plant viruses and is a heavy metal hyperaccumulator. However, little is known about the regulation of biotic and abiotic stress responses in this non-model plant. To investigate the control of miRNAs in gene expression, we sequenced the small RNA transcriptome of pokeweed treated with jasmonic acid (JA), a hormone that mediates pathogen defense and stress tolerance. We predicted 145 miRNAs responsive to JA, most of which were unique to pokeweed. These miRNAs were low in abundance and condition-specific, with discrete expression change. Integration of paired mRNA-Seq expression data enabled us to identify correlated, novel JA-responsive targets that mediate hormone biosynthesis, signal transduction, and pathogen defense. The expression of approximately half the pairs was positively correlated, an uncommon finding that we functionally validated by mRNA cleavage. Importantly, we report that a pokeweed specific miRNA targets the transcript of *OPR3*, novel evidence that a miRNA regulates a JA biosynthesis enzyme. This first large-scale small RNA study of a Phytolaccaceae family member shows that miRNA-mediated control is a significant component of the JA response, associated with widespread changes in expression of genes required for stress adaptation.

## 3.2 Introduction

The American pokeweed, *Phytolacca americana*, is a non-model plant with promising applications in agriculture. Pokeweed synthesizes pokeweed antiviral protein (PAP), a ribosome inactivating protein with RNA *N*-glycosidase activity (Endo et al., 1988). Several mutants of PAP have been generated and expressed heterologously in transgenic plants to impart novel antiviral and antifungal properties (Dai et al., 2003; Wang et al., 1998; Zoubenko et al., 1997, 2000). Additionally, pokeweed is a heavy metal hyperaccumulator with a potential role in phytoremediation (Liu et al., 2010; Peng et al., 2008; Zhao et al., 2011).

We recently assembled and annotated the pokeweed mRNA transcriptome and reported its regulation by jasmonic acid (JA; Neller et al., 2016). Jasmonates are well-characterized plant signal molecules of the oxylipin family that are involved in abiotic and biotic stress responses, as well as growth and development. JA mediates tolerance to drought, salt, heat and cold stresses, resistance to pathogens and insects, and induces embryogenesis, flowering, and senescence (reviewed in Wasternack and Feussner, 2017). JA is also implicated in tripartite interactions between insect vectors, viruses, and plants (Sun et al., 2017). Interestingly, JA treatment reduces heavy metal-induced oxidative stress (Singh and Shah, 2014) and conversely, heavy metal exposure is associated with biosynthesis of a JA precursor (Foroughi et al., 2014). Given that pokeweed exhibits broad-spectrum pathogen resistance, is a heavy metal hyperaccumulator, and undergoes marked transcriptional reprograming upon JA treatment, we are interested in exploring the range and regulation of defense strategies in this non-model plant.

JA biosynthesis is described by the octadecanoid pathway (reviewed in Schaller and Stintzi, 2009). Briefly, alpha-linolenic acid released from plastid membranes is oxidized by a13-lipoxygenase (LOX), followed by coupled dehydration-cyclization by allene oxide synthase

(AOS) and allene oxide cyclase (AOC) to generate 12-oxophytodienoic acid (OPDA). This undergoes reduction by OPDA reductase 3 (OPR3) and is followed by three rounds of beta oxidation to form JA. The JA-amido synthetase JAR1 produces the bioactive form of JA, (+)-7-iso-JAIle (JA-Ile); this is perceived by the COI1-JAZ co-receptor and results in ubiquitin-mediated degradation of JAZ repressors and activation of several transcription factors (reviewed in Chini et al., 2016).

Here, we complement our previous transcriptome analysis with a paired RNA-Seq study investigating miRNA-associated regulation of the JA response. miRNAs are non-coding RNAs of ~20-24 nucleotides (nt) that regulate post-transcriptional gene expression (reviewed in Yu et al., 2017). miRNA biogenesis begins with transcription of the *MIR* gene by RNA polymerase II to yield a primary miRNA (pri-miRNA), containing a stem-loop region flanked by unstructured arms. In sequential steps, dicer-like 1 RNase (DCL1) excises the stem-loop to form the pre-miRNA, then produces a smaller duplex comprising the miRNA and its opposing strand, classically termed miRNA*. In the cytoplasm, the miRNA associates with argonaute 1 (AGO1) to establish the RNA-induced silencing complex (RISC). Plant miRNAs act primarily through target mRNA cleavage, owing to high miRNA/target sequence complementarity, although translational inhibition has also been observed. miRNAs have been implicated in numerous abiotic and biotic stress responses, such as those induced by temperature, salinity, drought, light, nutrient deficiencies, oxidative stress, and pathogens (reviewed in Islam et al., 2018; Khraiwesh et al., 2012; Li et al., 2017). Previous reports have characterized the effect of JA or its methyl ester (JA-Me) on miRNAs in *Arabidopsis*, *Nicotiana attenuata*, *Taxus chinensis*, *Lycoris aurea* and *Catharanthus roseus* (Bozorov et al., 2012; Qiu et al., 2009; Shen et al., 2017; Xu et al., 2016; Zhang et al., 2012). Control of JA biosynthesis through miR319-targeted *TCP* transcription factors has also been

established (Schommer et al., 2008; Zhao et al., 2015). However, despite the importance of JA in plant development and stress resistance, the expression of JA-responsive miRNAs and targets has not been correlated on a genome-wide scale.

Although a Phytolaccaceae family reference genome is unavailable, current bioinformatic tools support miRNA analysis based on an assembled transcriptome (Dai and Zhao, 2011; Yang and Li, 2011). Here, we identify JA-responsive miRNAs, their potential roles, and construct a miRNA/target interaction network that incorporates correlated expression data from biological replicates. Our results indicate that many defense-related genes are regulated by JA-induced miRNAs, including the JA biosynthesis gene *OPR3*. These novel targets have potential agricultural applications to improve resistance to pathogen and environmental stresses.

## 3.3 Material and methods

### Plant growth conditions and jasmonic acid treatment

Pokeweed plants at the 4-leaf stage of growth were sprayed with 5 mM JA suspended in 0.5% ethanol, or 0.5 % ethanol alone (control), and leaf tissue was harvested 24 hours post treatment (Neller et al., 2016). Through time-course analysis, we established previously that PAP is expressed maximally 24 hours following JA treatment (Klenov et al., 2016). This time point was chosen for both the mRNA transcriptome-wide analysis of pokeweed (Neller et al., 2016) and the current paired miRNA analysis, as we are interested in identifying defense genes co-expressed with PAP.

**Small RNA sequencing and data processing**

Total RNA was extracted from leaf tissue of pokeweed plants as reported in Neller et al. (2016), and the small RNA fraction was sequenced according to Klenov et al. (2016). Briefly, the 15-30 nt size RNA portion was gel purified and ligated to directional 5p and 3p primers. Following cDNA synthesis and PCR-amplification, the small RNA libraries were sequenced on a SOLiD 5500 XL machine. Six libraries were sequenced, with three biological replicates per treatment. Each replicate consisted of an equal amount of small RNA pooled from three independent plants. The raw sequencing reads were processed as described in Klenov et al. (2016). Non-coding RNA (rRNA, tRNA, snRNA and snoRNA) and mature miRNA sequences were annotated by mapping reads to the sugar beet genome (Dohm et al., 2014) and the plant miRNA database (Zhang et al., 2010; http://bioinformatics.cau.edu.cn/PMRD/). Mapping was performed with Bowtie software (v. 1.0.1; Langmead et al., 2009) with up to three mismatches allowed in alignments.

**miRNA prediction and annotation**

miRNA prediction was performed with miRDeep-P (v. 1.3; Yang and Li, 2011), a modified version of the miRDeep algorithm (Friedländer et al., 2008) that incorporates an adapted scoring system for plant miRNA biogenesis. Non-annotated small RNA sequences as well as those annotated as mature miRNAs were used as input for miRNA prediction. Small RNA sequences were aligned to the pokeweed mRNA transcriptome (Neller et al., 2016), and only perfect, full-length alignments were retained. For each alignment, a potential precursor sequence of 250 nt in length was excised, and its secondary structure and minimum free energy were determined with RNAfold from the ViennaRNA package (v. 2.1.7; Lorenz et al., 2011). Mature miRNAs were predicted based on compatibility of the small RNA with the structure of its precursor. As per

miRDeep-P notation, if both 5p and 3p miRNA reads were detected then the most abundant strand was labelled as the mature miRNA. The less abundant strand was defined as the star sequence. Precursor structures were visualized with The UEA Small RNA Workbench (Stocks et al., 2012). Following miRNA prediction, BLASTn-short (Altschul et al., 1997) was used to identify conserved sequences ($E \leq 0.001$) against all mature sequences from the plant miRNA database. Remaining, non-annotated sequences were considered unique to pokeweed.

**miRNA differential expression analysis**

Raw sequencing counts of predicted miRNAs were used for differential expression analysis with the Bioconductor package EdgeR (Robinson et al., 2010) in the statistical program R (R Development Core Team, 2008). Prior to analysis, any miRNA with an abundance of less than 1 read per million (RPM) in each group of replicates was removed. Subsequently, libraries were normalized based on their total number of counts, and the common and tag-wise dispersions were estimated with default parameters. An Exact Test was conducted to detect miRNAs with significant differences in expression between control and JA-treated groups ($FDR < 0.05$). Each treatment group consisted of three biological replicates.

**Prediction of miRNA targets**

Predicted miRNAs were aligned to the pokeweed mRNA transcriptome and putative targets were identified using the Plant Small RNA Target Analysis Server with default parameters (Dai and Zhao, 2011; http://plantgrn.noble.org/psRNATarget/). This program predicts miRNA/target pairs based on reverse complementary matching and target site accessibility, incorporating plant-specific features of target recognition.

**Correlation of miRNA/target expression**

A paired design was employed for small RNA and mRNA sequencing; that is, both small RNA and mRNA fractions were sequenced for each biological replicate of each treatment group. The Pearson correlation coefficient (PCC) of each miRNA/target pair was calculated from normalized expression data in RPM and fragments per kilobase per transcript per million mapped reads (FPKM), respectively. Transcript quantification and differential expression results of miRNA targets were obtained from Neller et al. (2016).

**Functional analysis of miRNA targets**

Annotation information of miRNA targets was obtained from Neller et al. (2016). Gene ontology (GO) term distribution and enrichment analysis were conducted with Blast2GO software (Conesa et al., 2005). KOBAS 2.0 (Xie et al., 2011) was used to identify enriched KEGG pathways. For enrichment analysis, the test set comprised targets with highly correlated miRNAs (PCC > |0.8|) for which both the target and miRNA were differentially expressed (FDR < 0.05), and the reference set consisted of all miRNA targets.

**Construction of a miRNA/target interaction network**

Cytoscape software (v. 3.3.0; Shannon et al., 2003) was used to generate a network of highly correlated miRNA/target pairs (PCC > |0.8|) for which both members were differentially expressed (FDR < 0.05).

**Quantitative reverse transcription PCR (qRT-PCR)**

Reverse transcription was performed by combining 250 ng of total pokeweed RNA with 1 µL of 10 µM miRNA-specific stem-loop reverse primer and 1 µL of 10 mM dNTPs in a total volume of 13 µL. The solution was incubated at 65C for 10 minutes before cooling on ice for 3 minutes. To the mixture was added 4 µL of 5X First Strand Buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 1.38 µL of $dH_2O$, 1 µL of 0.1M DTT, 20 units of Murine RNase Inhibitor, and 25 units of Superscript III reverse transcriptase (Invitrogen). The reaction was incubated at 42C for 1 hour and heat inactivated at 70C for 20 minutes.

Following reverse transcription, 6 µL of cDNA was combined with 2 µL of 10 µM forward primer, 2 µL of 10 µM reverse primer, 23 µL of $dH_2O$, and 33 µL of 2X Sybr Green Mastermix (ABM). Each reaction was divided into three technical replicates and analyzed in a Qiagen Rotor-gene-Q real time PCR cycler. Ct values were calculated with the ΔΔCt relative quantification method. miR156 was used as an internal control, as it showed stable expression based on RNA-Seq results.

**5' rapid amplification of cDNA ends (RACE)**

Polyadenylated RNA was isolated with the NEBNext Poly(A) mRNA Magnetic Isolation Module. The method of 5' RACE was based on the Generacer 5' RACE kit, omitting phosphatase and decapping steps. To relax secondary structure, 250 ng of pokeweed mRNA and 250 ng of RNA adapter were combined in a total volume of 10 µL and incubated at 65C for 5 minutes. After cooling on ice for two minutes, 2 µL of 10X T4 RNA Ligase Buffer (500 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$, 10 mM DTT), 2 µL of 10 mM ATP, and 10 units of T4 RNA ligase were added and brought to a final volume of 20 µL. The ligation mixture was incubated at 37C for 1 hour,

extracted once with acidic phenol/chloroform, precipitated with 5 µg of linear acrylamide, and re-suspended in RNA storage buffer (22.5 mM DTT, 1 mM sodium citrate pH 6.4). Reverse transcription was performed as described above with poly-d(T) adapter primers, followed by one round of non-specific amplification with forward and reverse adapter primers. To isolate cleavage sites, 1 µL of cDNA was used as template for PCR with adapter forward primers and gene specific reverse primers 3' of the putative cleavage site. Following one round of nested PCR, the band of interest was gel purified, cloned, and sequenced. Five cDNA clones were sequenced per validation.

## 3.4 Results

### Overview of the pokeweed small RNA transcriptome

A summary of our study is provided in **Supplementary Fig. S1**. Following quality control processing, a total of 57,171,256 clean reads were obtained from control and JA-treated libraries (**Table 8**). To identify conserved miRNAs in pokeweed, reads were aligned to mature sequences from the plant miRNA database, with up to three allowable mismatches in alignments. Conserved miRNAs comprised a small percentage (< 0.5%) of reads, representing 1,381 unique sequences. Other classes of conserved, non-coding RNA in pokeweed (rRNA, tRNA, snRNA and snoRNA) were annotated by mapping reads to the sugar beet genome. Sugar beet was selected as a reference since it is the closest related species with a sequenced genome. At most, the combination of these remaining non-coding RNA species accounted for 14% of reads. Mapping rates to the pokeweed mRNA transcriptome averaged 18% and 23% for control and JA libraries, respectively. Although relatively low, these rates agree with the finding that most small RNAs were 24-mer heterochromatic siRNAs (59% and 55% for control and JA-treated libraries, respectively; **Fig. S2**). Small RNAs of this class are derived from non-polyadenylated precursors, so these reads were not

**Table 8. Annotation of small RNA sequences in pokeweed**.

| Category | C1 | C2 | C3 | JA1 | JA2 | JA3 |
|---|---|---|---|---|---|---|
| **clean reads** | 12,794,660 | 8,560,268 | 8,109,969 | 9,697,656 | 7,997,366 | 10,011,337 |
| | **(100%)** | **(100%)** | **(100%)** | **(100%)** | **(100%)** | **(100%)** |
| **miRNA** | 28,149 | 28,249 | 28,385 | 32,002 | 25,592 | 34,039 |
| | **(0.22%)** | **(0.33%)** | **(0.35%)** | **(0.33%)** | **(0.32%)** | **(0.34%)** |
| **rRNA** | 220,069 | 111,283 | 97,320 | 180,376 | 110,364 | 218,247 |
| | **(1.72%)** | **(1.30%)** | **(1.20%)** | **(1.86%)** | **(1.38%)** | **(2.18%)** |
| **tRNA** | 319,867 | 215,719 | 223,835 | 662,350 | 570,212 | 1,098,243 |
| | **(2.50%)** | **(2.52%)** | **(2.76%)** | **(6.83%)** | **(7.13%)** | **(10.97%)** |
| **snRNA** | 6,141 | 3,339 | 2,436 | 5,830 | 4,798 | 6,028 |
| | **(0.05%)** | **(0.04%)** | **(0.03%)** | **(0.06%)** | **(0.06%)** | **(0.06%)** |
| **snoRNA** | 6,653 | 3,510 | 2,430 | 5,808 | 5,598 | 5,986 |
| | **(0.05%)** | **(0.04%)** | **(0.03%)** | **(0.06%)** | **(0.07%)** | **(0.06%)** |
| **mRNA transcriptome** | 2,566,606 | 1,687,229 | 1,428,977 | 2,175,184 | 1,848,191 | 2,474,803 |
| | **(20.06%)** | **(19.71%)** | **(17.62%)** | **(22.43%)** | **(23.11%)** | **(24.72%)** |
| **unannot.** | 9,647,174 | 6,510,940 | 6,326,586 | 6,636,106 | 5,432,611 | 6,173,991 |
| | **(75.40%)** | **(76.06%)** | **(78.01%)** | **(68.43%)** | **(67.93%)** | **(61.67%)** |

*The small RNA fraction was sequenced from leaf tissue of 4-leaf plants treated with ethanol (C, control) or jasmonic acid (JA). Clean reads were mapped to the sugar beet genome and mature miRNA sequences from the plant miRNA database, with three allowable mismatches, and to the pokeweed mRNA transcriptome, with perfect alignment. Three biological replicates were sequenced per treatment.*

expected to align to the mRNA transcriptome. Taken together, most of the pokeweed small RNA transcriptome remains un-annotated when compared with sequences from other plants. This indicates that the Phytolaccaceae family represents a source of novel small RNAs and corresponding targets.

**miRNA prediction and differential expression analysis**

Pokeweed miRNAs were predicted computationally with miRDeep-P, a plant-specific program that incorporates principles of miRNA biogenesis. Briefly, unannotated reads and those annotated as mature miRNA sequences were aligned to the pokeweed mRNA transcriptome to identify likely miRNA precursors. For each alignment, a flanking region of mRNA was extracted and its secondary structure predicted. Mature miRNAs were identified based on the likelihood that a particular small RNA sequence originated from the biological processing of a miRNA precursor. This strategy led to the prediction of 582 miRNAs in pokeweed; their sequences are provided in **Supplementary Table S1**. To identify which of these miRNAs were conserved, a BLASTn-short search was conducted against mature sequences from the plant miRNA database (E value $\leq 0.001$). A total of 24 conserved miRNAs were annotated (**Table 9**). Among these, the most abundant were miR156, miR169, miR535, and miR396.

The majority of predicted miRNAs were 21-nt, as expected (**Figure 11A**); however, this pool had low sequence diversity, as many 21-nt reads had the same sequence (**Figure 11B**). Conversely, non-canonical 24-nt miRNAs had high sequence diversity and likely targeted numerous different genes, even though they comprised a small fraction of the miRNA pool. These trends were consistent following differential expression analysis to identify JA-responsive

**Table 9. Conserved miRNAs in pokeweed.**

| miRNA | Sequence (5'-3') | Abundance (RPM) |
|---|---|---|
| 156 | UGACAGAAGAGAGUGAGCAC | 2836.34 |
| 157 | GCUCUCUAUGCUUCUGUCAUC | 9.29 |
| 159 | GAUCAUGUGGUAGCUUCACC | 203.98 |
| 162 | UCGAUAAACCUCUGCAUCCAG | 121.20 |
| 164 | UGGAGAAGCAGGGCACGUGCA | 531.04 |
| 166 | UCGGACCAGGCUUCAUUCCUC | 16.09 |
| 166 | CCGGACCAGGCUUCAUUCCCC | 14.78 |
| 166 | UCGGACCAGGCUUCAUUCCCC | 11.21 |
| 167 | UGAAGCUGCCAGCAUGAUCUG | 182.36 |
| 168 | UCGCUUGGUGCAGGUCGGGAA | 94.33 |
| 169 | CAGCCAAGGAUGACUUGCCGG | 2126.80 |
| 169 | UAGCCAAGGAUGACUUGCCUG | 131.25 |
| 171 | UGAUUGAGCCGUGCCAAUAUC | 412.17 |
| 171 | UGAUUGAGCCGCGCCAAUAUC | 193.00 |
| 172 | CGAAUCUUGAUGAUGCUGCAU | 69.79 |
| 319 | UUUGGAUUGAAGGGAGCUC | 9.66 |
| 394 | UUUGGCAUUCUGUCUACCUCC | 2.57 |
| 395 | CUGAAGUGUUUGGGGGAACUC | 7.24 |
| 396 | UUCCACAGCUUUCUUGAACUG | 1897.91 |
| 396 | UUCCACAGCUUUCUUGAACUU | 885.08 |
| 403 | UUAGAUUCACGCACAAACUCG | 141.05 |
| 535 | UGACGAUGAGAGAGAGCACGC | 1991.70 |
| 3954 | UUAGACAGAGAAAUCACGGUUG | 49.85 |
| 5203 | AGUGACAGAUAUUAUGGACCGGAG | 0.26 |

*Following miRNA prediction, conserved sequences were identified through BLASTn-short alignment ($E \leq 0.001$) against the plant miRNA database. The average abundance of each miRNA is provided in reads per million (RPM).*

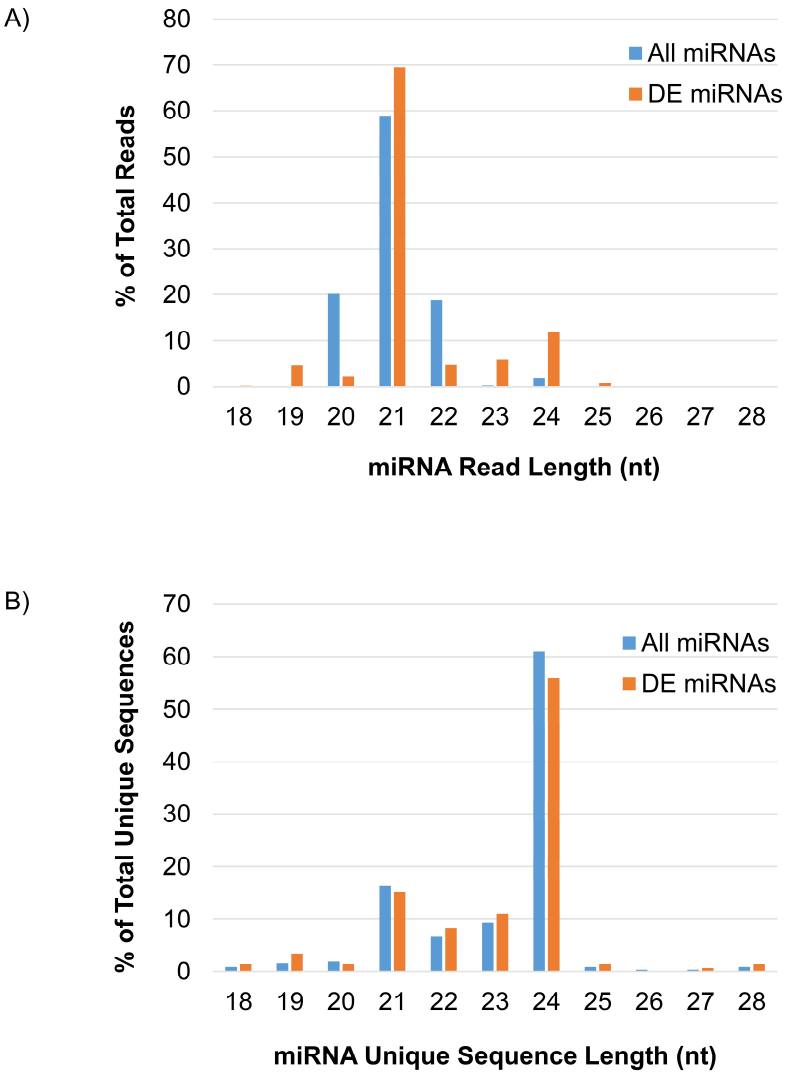



**Figure 11. The length distribution of predicted miRNAs in pokeweed. A)** reads; **B)** unique sequences. DE, differentially expressed; nt, nucleotides.

miRNAs. That is, the 24-nt miRNAs contained the greatest number of unique sequences in both the total miRNA and JA-responsive miRNA pools.

Of 582 miRNAs, 145 showed significant changes in abundance with JA treatment, relative to control (FDR < 0.05; **Figure 12A** and **Supplementary Table S2**). Among these were two conserved miRNAs, miR172 and miR395, with $\log_2$ fold changes of -13.4 and 2.30, respectively. The numbers of up- and down-regulated miRNAs were equal, with 73 and 72 sequences, respectively. Interestingly, although a majority, 443 of the 582 miRNAs, were present in both control and JA samples (**Figure 12B**), most differentially expressed miRNAs were condition-specific (**Figure 12C).** That is, most were detected in either all control replicates or all treated replicates; only 16 miRNAs (11% of those differentially expressed) were found in all samples. Furthermore, most JA-responsive miRNAs were of low abundance and displayed high fold changes within a narrow range (**Figure 12D**). We observed a symmetrical distribution of miRNA fold changes (**Figure 12E**). Of all differentially expressed miRNAs, those that decreased in abundance with JA had an average $\log_2$FC and standard deviation of -5.13 ± 1.38, while those that increased with JA had values of 4.92 ± 1.07. Therefore, JA treatment alters sequence diversity within the miRNA pool, and discrete fold changes in expression of JA-responsive miRNAs may be required for biological effect.

**Identification of JA-responsive miRNA/target pairs**

We used the Plant Small RNA Target Analysis Server to identify pokeweed mRNA transcripts having high sequence complementarity with the 582 predicted miRNAs, which resulted in a total of 25,931 miRNA/target pairs. To identify the most biologically relevant interactions, the list was filtered to include only JA-responsive miRNAs and targets (FDR < 0.05), resulting in

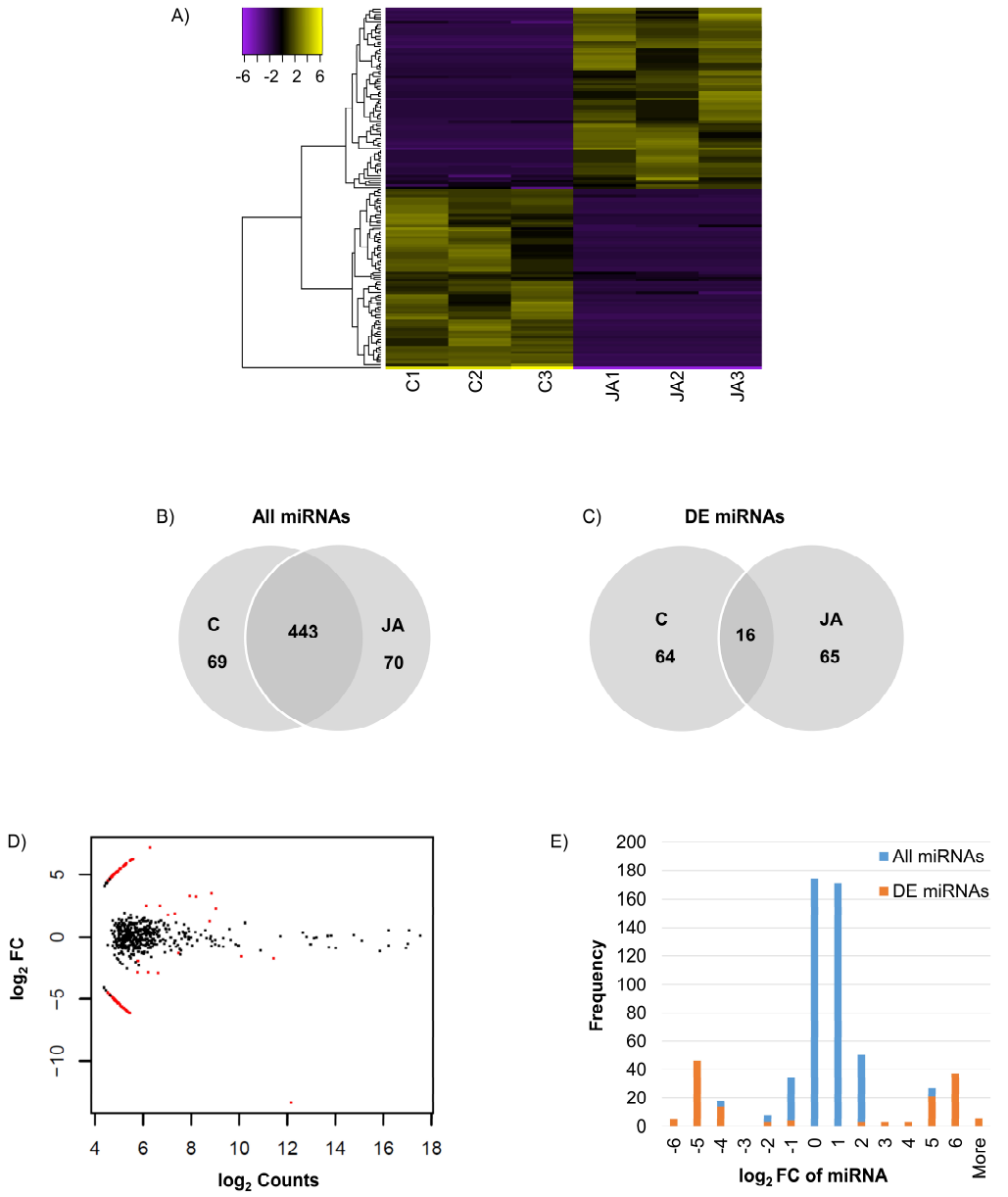



**Figure 12. Identification of JA-responsive miRNAs in pokeweed. A)** Heat map of expression values ($log_2$RPM, median-centered) of differentially expressed miRNAs (FDR < 0.05). **B)** and **C)** depict treatment-specific expression patterns of all miRNAs and differentially expressed miRNAs, respectively. **D)** MA plot indicating all differentially expressed miRNAs (red). **E)** Histogram of miRNA fold change; x-axis values indicate upper bin numbers. DE, differentially expressed; FC, fold change.

a subset of 516 miRNA/target pairs. That is, for each pair, both miRNA and target were differentially expressed with JA. The rationale for limiting our analysis in this manner was to include only interactions that could potentially mediate a functional response in the plant through the alteration of mRNA abundance, given that plant miRNAs act predominately through target mRNA cleavage.

The expression change of each miRNA and target forming the 516 JA-responsive pairs is shown in **Figure 13A**. The number of negative and positive miRNA/target expression correlations was similar, with 274 pairs showing opposite JA-induced changes (i.e. miRNA ↑ and target ↓ or vice versa) and 242 pairs with parallel changes (i.e. miRNA and target both ↑ or ↓). Most strikingly, while miRNA expression change was limited to a narrow range, target change was much more variable. Interestingly, plotting all miRNA/target pairs revealed that the non-random distribution was inherent in the overall miRNA population, as shown by trimodal clustering (**Figure 13B**). These results indicate that JA treatment induces a certain threshold of miRNA expression change, for those miRNAs that are JA-responsive, suggesting an optimal miRNA fold change for effective JA control.

This RNA-Seq study incorporated paired small RNA and mRNA samples; that is, each RNA sample was sequenced to quantify both small RNA and mRNA levels. This paired design, combined with the use of independent biological replicates, enabled the identification of miRNA/target pairs with highly correlated expression. The PCC was calculated for each of the 516 miRNA/target pairs described above, and the distribution of correlations is shown in **Supplementary Fig. S3**. There was a clear enrichment in moderate and high correlations; this was expected as differential expression analysis and sequence complementarity filters had already been applied to retain the most likely miRNA/target candidates. Specifically, we were interested in

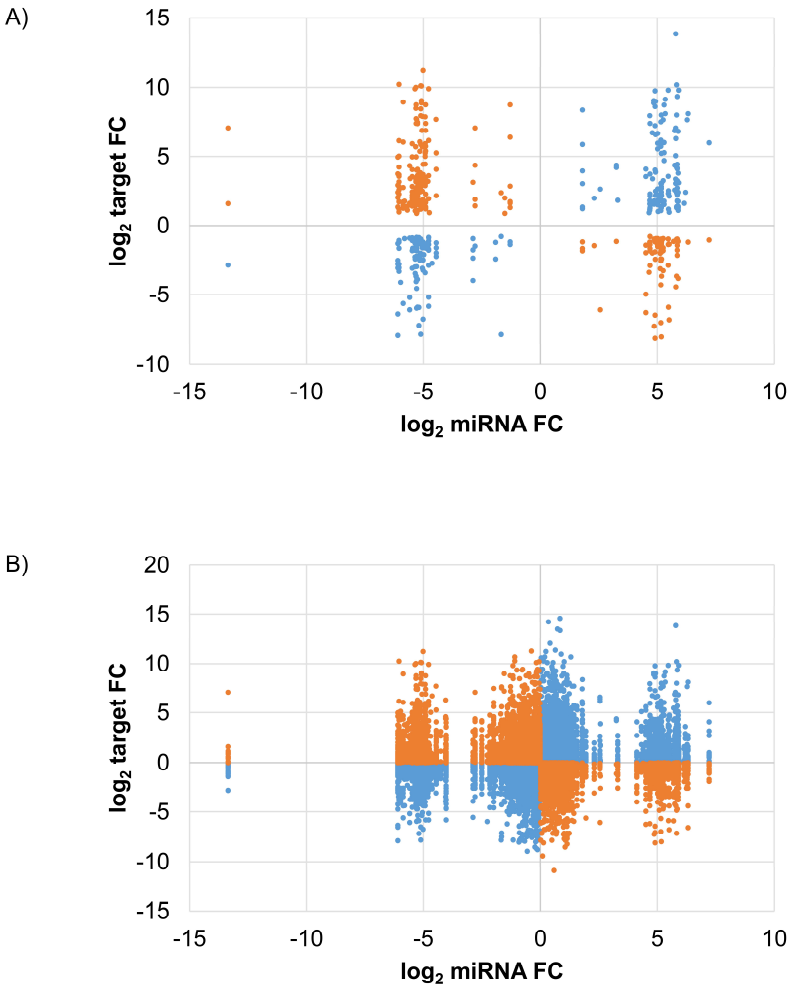



**Figure 13. Comparison of miRNA and target expression changes.** The fold change (FC) of each miRNA was plotted against that of its predicted target. Blue and orange indicate positive and negative correlations, respectively, of miRNA/target expression changes. **A)** Differentially expressed miRNAs and targets (FDR < 0.05). **B)** All miRNAs and targets.

highly correlated pairs, defined as those with PCC > |0.8|. This cut-off reduced the number of pairs from 516 to 171, comprising 97 positively correlated pairs and 74 negatively correlated ones. These 171 pairs are most biologically relevant since the expression of each miRNA is associated with the expression of its target.

**Functional analysis of highly correlated miRNA/target pairs**

Following the identification of highly correlated miRNA/target pairs, our next goal was to functionally characterize them. Target annotations were obtained from our previous study of the pokeweed mRNA transcriptome (**Supplementary Table S3**). Based on KEGG pathway enrichment, the top pathways were associated with plant stress and defense responses, including "peroxisome", "alpha-Linolenic acid metabolism", and "plant-pathogen interaction" (**Table 10**). The corresponding targets mapping to each enriched pathway revealed key genes that may be regulated by miRNAs. Notably, these included two genes associated with JA biosynthesis (*OPR3* and *AOC3*) as well as an integrator of JA and ethylene signalling (Ethylene-responsive transcription factor 1B, *ERF1B*). Although miRNA-mediated control of transcription factors involved in JA biosynthesis has been reported, our results indicate that miRNAs directly target genes of the JA biosynthesis pathway.

Since pokeweed is a non-model plant with limited resources for annotation, some targets could not be mapped to GO terms or KEGG pathways, or their functional associations lacked substantial detail. To account for this, we manually assessed the 171 highly correlated miRNA/target pairs and cross-referenced the literature to find genes associated with plant stress and defense responses. The results are summarized in **Table 11**, revealing novel targets involved in hormone signalling, hypersensitive response, lignin biosynthesis, and oxidative stress response.

**Table 10. Top pathways involving highly correlated, JA-responsive miRNA/target pairs.**

| Pathway | FDR | Annotated Genes |
|---|---|---|
| Peroxisome | 0.077 | 2-hydroxyacyl-CoA lyase<br>Peroxisomal nicotinamide adenine dinucleotide carrier<br>Isocitrate dehydrogenase |
| alpha-Linolenic acid metabolism | 0.077 | Allene oxide cyclase 3, chloroplastic<br>12-oxophytodienoate reductase 3 |
| Plant-pathogen interaction | 0.077 | Calcium-dependent protein kinase 16<br>Calmodulin-like protein 1 |
| Glutathione metabolism | 0.12 | Isocitrate dehydrogenase [NADP]<br>Glutathione S-transferase U9 |
| Plant hormone signal transduction | 0.13 | Ethylene-responsive transcription factor 1B<br>Auxin response factor 1<br>Two-component response regulator ARR9 |

*KEGG pathway enrichment was performed relative to all miRNA/target pairs, which served as the reference set. The specific genes annotated in each enriched pathway are indicated.*

**Table 11. Identification of highly correlated, JA-responsive miRNA/target pairs implicated in plant stress and defense responses.**

| miRNA | | Target | | | |
|---|---|---|---|---|---|
| **ID** | **log$_2$FC** | **ID** | **log$_2$FC** | **Name** | **Function** |
| 204283_x13 | -5.11 | c113292_g1_i1 | 2.82 | Allene oxide cyclase 3, chloroplastic | JA biosynthesis (Ziegler et al., 2000) |
| 98462_x2365 | -1.69 | c19572_g1_i1 | 2.36 | 12-oxophytodienoate reductase 3 | JA biosynthesis (Schaller et al., 2000) |
| 419489_x11 | 4.92 | c37552_g1_i1 | 9.01 | Ethylene-responsive transcription factor 1B | JA signalling (Lorenzo et al., 2003) |
| 404574_x21 | 5.83 | c58992_g2_i1 | 8.02 | NAC domain-containing protein 72 | abiotic stress-associated transcription factor (Tran et al., 2004) |
| 499371_x14 | 5.29 | c58861_g12_i1 | 1.51 | Auxin response factor 1 | transcription repressor, auxin response (Ellis et al., 2005) |
| 297830_x3990 | -13.36 | c40819_g1_i1 | -2.81 | Two-component response regulator ARR9 | negative regulator of cytokinin signalling (To, 2004) |
| 430760_x13 | 5.15 | c47336_g1_i1 | 6.17 | Zinc finger protein ZAT10 | transcription repressor, abiotic stress response (Mittler et al., 2006) |
| 144454_x98 | 1.80 | c17159_g1_i1 | 5.88 | Transcription factor bHLH25 | JA-induced transcription factor (Heim et al., 2003) |
| 319549_x8 | -4.46 | c60508_g2_i1 | 2.14 | TGACG-sequence-specific DNA-binding protein TGA-2.1 | SA-induced transcription activator (Niggeweg et al., 2000) |
| 225512_x26 | -6.05 | c52120_g1_i1 | 2.58 | Probable protein phosphatase 2C 73 | negative regulator of ABA signalling (Umezawa et al., 2009) |
| 66102_x232 | 3.25 | c33290_g1_i1 | 4.22 | U-box domain-containing protein 19 | protein ubiquitination; negative regulator of ABA and drought responses (Liu et al., 2011) |
| 513239_x9 | 4.68 | c42333_g1_i1 | 1.72 | F-box protein FBW2 | protein ubiquitination; response to ABA (Earley et al., 2010) |
| 110629_x73 | -2.89 | c54352_g2_i1 | -1.73 | Glucan endo-1,3-beta-glucosidase 1 | anti-microbial activity; viral pathogenesis (Beffa et al., 1996) |
| 110629_x73 | -2.89 | c13510_g1_i1 | 3.13 | Polygalacturonase inhibitor | anti-microbial activity (Kalunke et al., 2015) |
| 456330_x11 | 3.90 | c80075_g1_i1 | 4.92 | Snakin-2 | anti-microbial activity (Berrocal-Lobo et al., 2002) |
| 443361_x21 | 5.89 | c58046_g10_i1 | -3.80 | Probable cinnamyl alcohol dehydrogenase 6 | lignin biosynthesis (Sibout, 2005) |
| 443361_x21 | 5.89 | c23015_g1_i1 | 1.73 | Dirigent protein 21 | lignin biosynthesis (Davin and Lewis, 2000) |
| 319549_x8 | -4.46 | c17278_g1_i2 | 5.28 | Uclacyanin-2 | lignin biosynthesis (Nersissian et al., 1998) |
| 419489_x11 | 4.92 | c57722_g7_i1 | 2.17 | Glutathione S-transferase U9 | cellular detoxification (Wagner et al., 2002) |
| 478663_x13 | 5.18 | c54687_g4_i1 | 3.06 | Isocitrate dehydrogenase [NADP] | response to oxidative stress (Valderrama et al., 2006) |
| 1213_x8 | 4.50 | c56365_g1_i1 | -1.39 | Proline-rich receptor-like protein kinase PERK1 | hypersensitive response (Silva and Goring, 2002) |
| 485379_x20 | 5.78 | c59820_g2_i8 | 3.04 | Calcium-dependent protein kinase 16 | hypersensitive response (Coca and San Segundo, 2010) |
| 513239_x9 | 4.68 | c48318_g1_i3 | -2.85 | Hypersensitive-induced response protein 1 | hypersensitive response (Qi et al., 2011) |

*Log$_2$FC indicates fold change of miRNA and target levels in JA-treated libraries relative to control libraries.*

Importantly, some miRNAs in **Table 11** have more than one target, such as 110629_x73, which is predicted to regulate two genes encoding anti-microbial proteins, and 419489_x11, which targets genes involved in both JA signalling and cellular detoxification. Taken together, it is evident that targets of highly correlated miRNAs contribute to various aspects of biotic and abiotic stress responses.

We constructed an interaction network to visualize the 171 highly correlated, differentially expressed miRNA/target pairs (**Supplementary Fig. S4**). The network includes 67 unique miRNAs, the majority of which (40) had more than one target. Major hubs are located towards the top of the network; these are cases where a single miRNA has many targets, and different hubs are often connected by a common target. Enrichment analysis was conducted to identify overrepresented GO terms associated with major hubs; however, no significant terms were found. This suggests that though individual miRNA/target pairs are involved in plant defense, these miRNAs also control different mRNAs with diverse functions, rather than all targets of a single miRNA being involved in a single response.

Precursor structures of the 67 miRNAs are provided in **Supplementary Fig. S5**. Although variability was evident in terms of overall precursor shape and positioning of mature sequence, the mature miRNA tended to be in the middle or upper stem of a stable hairpin structure. We did note a few unexpected cases wherein the mature miRNA was mostly or completely contained within a non-hairpin loop; these cases may have arisen from limitations in the RNA folding and/or miRNA prediction programs.

**Validation of miRNA differential expression and functional activity**

To validate RNA-Seq results, we measured the expression of eight miRNAs (98462_x2365, 297830_x3990, 66102_x232, 110629_x73, 114897_x905, 144454_x98, 62255_x144, 50891_x48) by qRT-PCR from control and JA-treated plants (**Figure 14**). These miRNAs were selected to represent a range of fold changes ($9.5x10^{-5}$ to 9.5) and abundances (0.4 to 44 RPM), and all had a highly correlated target that was functionally annotated through BLASTx. The $R^2$ correlation value was 0.88, indicating high correspondence between the two methods of differential expression analysis. The fold-change of transcript levels quantified with RNA-Seq were higher than those quantified with qRT-PCR, as observed previously (Baran-Gale et al., 2015). This difference is likely due to lower amplification efficiency of stem-loop primers used in qRT-PCR, compared to adapter-specific primers used for RNA-Seq. To verify miRNA-induced cleavage of a specific mRNA, four targets were validated by 5'RACE: c56431_g2_i2 (squamosa promoter-binding-like protein 2*, SPL2*), c59355_g7_i1 (thymidine kinase*, TK*), c19572_g1_i1 (*OPR3)*, and c17159_g1_i1 (transcription factor bHLH25*, BHLH25)* (**Figure 15A**). The *SPL2*/miR156 pair was selected as a positive control, as it is a well-established interaction. mRNA cleavage was detected at the expected site, between the $10^{th}$ and $11^{th}$ nucleotide of the miRNA binding site, in all cases except for *OPR3*, where cleavage occurred between the $11^{th}$ and $12^{th}$ nucleotide. *OPR3* and *BHLH25* were targeted by pokeweed-specific miRNAs (98462_x2365 and 144454_x98, respectively) and represent two confirmed interactions from **Table 11**; both miRNAs were also validated through qRT-PCR (above). *TK* was identified unexpectedly as a novel target of the well-conserved miR156 and was significantly JA-responsive based on mRNA-Seq results. Precursor structures of miR156 and the two novel miRNAs are shown in **Figure 15B**.

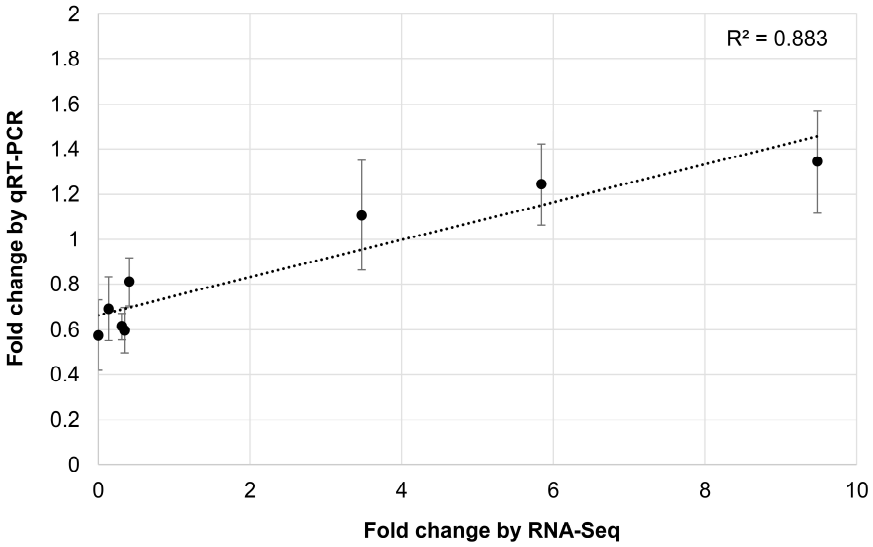
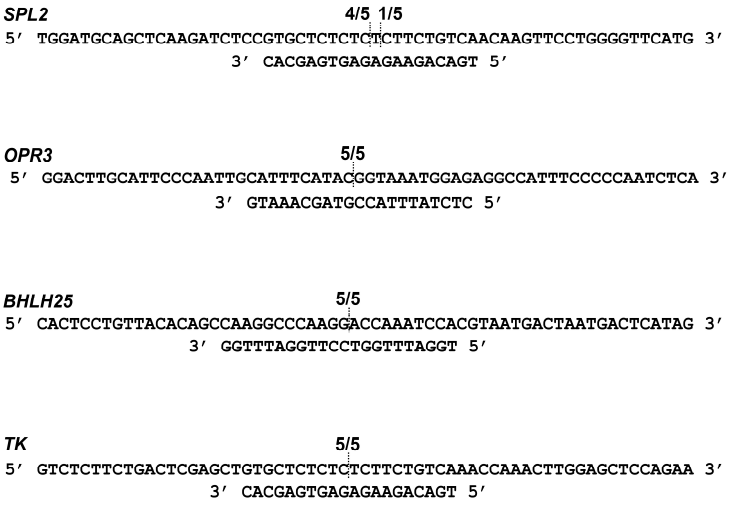



**Figure 14. Validation of JA-responsive miRNAs.** The correlation of JA-induced expression changes obtained from RNA-Seq and qRT-PCR is shown for 8 miRNAs. Results for qRT-PCR represent the mean from two or three independent biological replicates for each miRNA. Bars indicate the standard error associated with qRT-PCR.

A)




B)

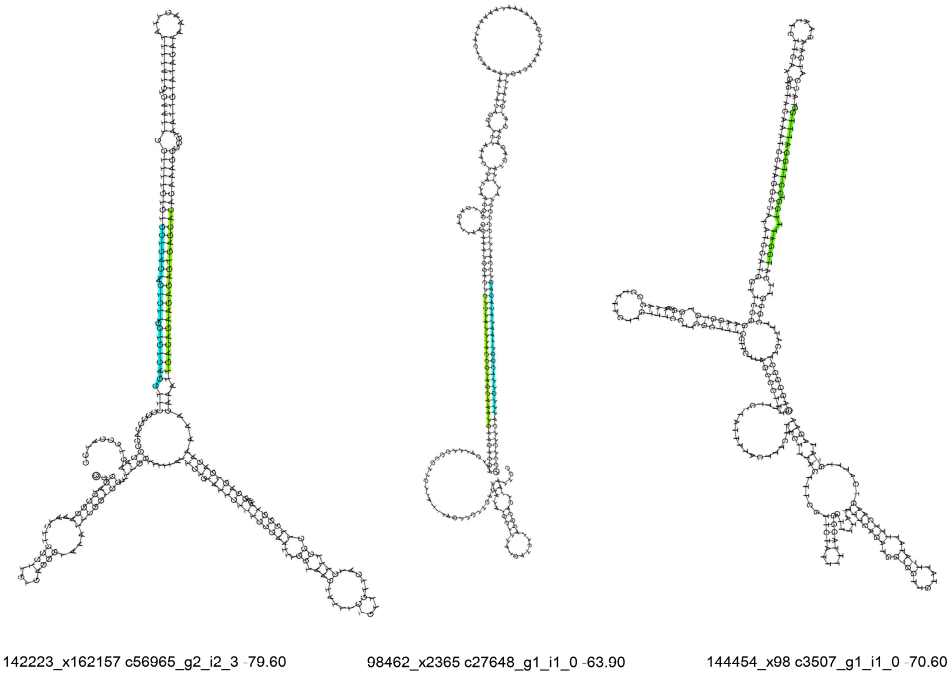



**Figure 15. Validation of miRNA target cleavage. A)** miRNA-induced cleavage of mRNA targets was validated through 5'RACE. The upper sequence indicates the mRNA target and the lower sequence is its miRNA. Fractions indicate cleavage sites from individual cDNA clones. From top to bottom, miRNAs are: miR156, miR98462_x2365, miR144454_x98, and miR156. **B)** Precursor structures are shown for each miRNA. The mature miRNA sequence is indicated in green. If a miRNA star strand was detected, it was indicated in blue. For each structure, the corresponding miRNA ID, precursor mRNA ID, and minimum free energy are provided.

Both miR156 and miR 98462_x2365 had detectable star sequences while 144454_x98 did not, likely owing to its lower abundance. All three precursors adopted stable hairpin formations.

### 3.5 Discussion

We have performed a large-scale miRNA analysis of pokeweed, with emphasis on JA-induced genes that contribute to biotic and abiotic stress tolerance. A comprehensive resource of 57,171,256 high-quality small RNA reads has been generated, representing the first available small RNA transcriptome of any Phytolaccaceae family member. Through miRNA prediction and annotation, we identified conserved and novel miRNAs in pokeweed. Differential expression analysis revealed that the majority of JA-responsive miRNAs were low in abundance, condition-specific, and exhibited a narrow range of expression change. Our paired RNA-Seq experimental design allowed the prediction of high-confidence miRNA targets and construction of an interaction network to visualize the global connection between miRNAs and their different targets.

#### Pokeweed miRNAs have low sequence conservation with those of other plants

Unexpectedly, less than 0.5% of pokeweed small RNA reads represented conserved miRNAs, based on comparison with available plant sequences. Compared with other primary studies of non-model plants, this number seems low. For example, in black pepper, 11% of small RNA reads were conserved miRNAs, with up to two mismatches in alignments (Asha et al., 2016). In *Vriesea carinata*, 16% of small RNA reads were conserved miRNAs, with only perfect alignments permitted (Guzman et al., 2013). Since this is the first miRNA study of any Phytolaccaceae member, low sequence conservation in pokeweed may be a result of species

specificity. Although the majority of plant miRNAs are species-specific (Chávez Montes et al., 2014), the low level of conservation in pokeweed is indicative of a relatively unique miRNA pool.

We performed a stringent miRNA annotation by requiring precursor evidence in addition to mature sequence conservation. Using this strategy, 24 conserved miRNAs were identified in pokeweed, and their expression levels were similar to those observed in other plants (Chávez Montes et al., 2014). Although 1,381 sequences were conserved, a miRNA precursor was predicted for only 1.6%. Others have reported similar findings. For example, in black pepper, 33,350 miRNA sequences were conserved but precursors were found for only 50, or 0.15% (Asha et al., 2016). Despite the common strategy of identifying miRNAs based on sequence conservation alone, our analysis indicates that this method may overestimate the number of true miRNAs.

We also observed non-canonical miRNAs that were 24-nt in length and had high sequence diversity relative to 21-mers. Similar findings have been reported by others (Guzman et al., 2013; Hackenberg et al., 2013; Shuai et al., 2013; Wan et al., 2012). Longer miRNAs tend to be species-specific and are reflective of newly evolved "proto-miRNAs" that undergo imprecise processing by DCL enzymes; interestingly, 24-nt miRNAs can enter the heterochromatic siRNA pathway to direct chromatin modifications of target genes (Axtell, 2013).

**JA treatment induces defined changes in the miRNA transcriptome**

Differential expression analysis revealed that JA-responsive miRNAs tended to be condition-specific and low in abundance, with narrowly clustered expression patterns. Condition specificity and low abundance may indicate that, at certain times, sequence diversity of the miRNA pool is more biologically relevant than miRNA abundance. For example, low abundance miRNAs may be early responders to stress. In support of this idea, Pandey et al. (2008) reported only 2.4%

overlap between the small RNA populations of untreated *Nicotiana attenuata* plants and those having undergone 45 minutes of insect elicitation, which strongly upregulated the JA response. Since low abundance miRNAs tend to be species-specific (Chávez Montes et al., 2014), analysis of corresponding targets may lead to the identification of key genes contributing to distinctive traits.

Our expression results suggest that a certain threshold of miRNA change is necessary to mediate an effective response to JA. Specifically, $\log_2$FC values of JA-responsive miRNAs clustered around +5 and -5. JA-associated thresholding is further supported by the finding that $\log_2$FC values of insect-induced small RNAs in *N. attenuata* ranged from +4 to -4 (Pandey et al., 2008). Clustering of miRNAs has also been reported. In cotton under salt stress, Peng et al. (2014) observed four discrete groups when the expression of salt-responsive miRNAs was plotted against that of targets; such clustering existed for both positively and negatively correlated pairs. These patterns may reflect co-regulated miRNA expression owing to control by common transcription factors, which are themselves differentially expressed.

Although we identified many JA-responsive miRNAs, only two were conserved with other plants: miR172 and miR395, which decreased and increased with JA, respectively. In contrast, others have reported several conserved, JA-responsive miRNAs (Bozorov et al., 2012; Qiu et al., 2009; Shen et al., 2017; Xu et al., 2016; Zhang et al., 2012). One reason for this discrepancy may be our stringent requirements for differential expression, owing to the availability of biological replicates. Other important differences between our work and that cited above include the use of plants deficient in JA biosynthesis, as well as the investigation of cell cultures rather than whole plants. Although miR172 was reported as JA-responsive by others, its magnitude and direction of change differed across studies. In contrast, miR395 has not been reported as JA-responsive; its

most well-established role to date is in sulfur assimilation (Matthewman et al., 2012). Interestingly, miR395 increased in *Arabidopsis* after heavy metal treatment and contributed to cadmium detoxification in *Brassica napus* (Jagadeeswaran et al., 2013; Zhang et al., 2013a). Given that heavy metal stress is associated with biosynthesis of a JA precursor (Foroughi et al., 2014), and JA treatment reduces heavy metal stress (Singh and Shah, 2014), there is a likely association between miR395 and JA.

**JA-responsive miRNAs are both positively and negatively correlated with targets**

To identify the top miRNA/target candidates, three factors were considered: differential expression, sequence complementarity, and correlation analysis. Pearson correlation is widely used to infer miRNA/target relationships (Muniategui et al., 2013). However, since high correlation does not necessarily indicate causation, we applied this filter after first considering more biologically relevant parameters. Although there is a tendency in literature to report only pairs with negatively correlated expression, we observed a relatively equal distribution of positive and negative correlations. This finding has been observed in other plant studies, reinforcing the concept that both dynamics are prevalent and worthy of consideration (Lopez-Gomollon et al., 2012; Peng et al., 2014; Wen et al., 2016). Positive correlations between the expression of a miRNA and its target can arise from miRNA-mediated spatial restriction of a target (Kawashima et al., 2009; Kidner and Martienssen, 2004; Levine et al., 2007; Nikovics et al., 2006). In *Arabidopsis* roots, a positive temporal and negative spatial correlation was found for the expression of miR395 and its target, sulphur transporter SULTR2;1. Authors hypothesized that under sulphur starvation, phloem-specific co-expression of miR395/SULTR2;1 restricts expression of SULTR2;1 to the xylem (Kawashima et al., 2009). As another example, the extent

of leaf serration in *Arabidopsis* is controlled by spatial restriction of MIR164A-targeted CUC2 (Nikovics et al., 2006).

Additionally, it is thought that miRNAs operate through two independent mechanisms that are not mutually exclusive: expression tuning, whereby they modify the mean target abundance, and expression buffering, in which they reduce the variance of target abundance around a preset mean (Wu et al., 2009). Negative miRNA/target expression correlations could arise from the first mechanism and positive correlations from the second. Expression buffering consists of negative feedback loops and incoherent feed forward loops (Tsang et al., 2007; Wu et al., 2009). Consider a miRNA, "miR" and its target gene, *"T*". As summarized by Wu et al., in a negative feedback loop, protein T is a positive regulator of miR, such that increased T results in increased miR, dampening oscillations in the expression of either gene. An example of this dynamic has been shown within the miRNA biogenesis pathway itself: miR162 targeting of *DCL1* (Xie et al., 2003). In an incoherent feed forward loop, consider an additional gene, *A*, whose product increases levels of both T and miR. Since A regulates the level of T, any change in the level of A would also affect that of T. However, variation in A is also passed to miR, which then buffers the fluctuation in T. An interesting example here is the maintenance of AGO1 homeostasis by 1) miR168 targeting; 2) transcriptional co-expression of miR168 and AGO1; and 3) preferential stabilization of miR168 by AGO1. This example combines two negative feedback loops (1 and 3), with an incoherent feed forward loop (2). Though positive correlations of miRNA and targets gene levels are not commonly described, examples exist of their biological relevance.

We validated the miRNA-induced cleavage of a positively correlated target, *BHLH25*. Little is known about the biological role of this transcription factor, but it is reportedly JA-responsive (Heim et al., 2003), and transgenic *Arabidopsis* plants overexpressing this gene have

altered root and shoot morphology and greater susceptibility to cyst nematode infection (Jin et al., 2011). Increased expression of the miRNA in tandem with *BHLH25* could buffer levels of this transcription factor during JA induction. For example, an increase in BHLH25 protein could increase the expression of its own miRNA directly (negative feedback loop), and/or an upstream transcription factor could simultaneously increase expression of both BHL25 and its miRNA (incoherent feed forward loop).

**JA-responsive miRNA/target pairs are associated with multiple levels of plant defense**

We identified putative miRNA targets involved in JA biosynthesis and signalling. Two genes encoding JA biosynthesis enzymes were identified, *AOC3* and *OPR3,* and the miRNA-induced cleavage of *OPR3* mRNA was verified by 5' RACE. *OPR3* and more recently, *OPR2*, were shown to be required for JA biosynthesis in *Arabidopsis* (Chini et al., 2018; Stintzi and Browse, 2000). Although miR319 targets *TCP* transcription factors and indirectly regulates JA biosynthesis (Schommer et al., 2008), this is the first reported miRNA-induced cleavage of a transcript encoding a JA biosynthesis enzyme. Others have predicted complementary miRNA/target pairs comprising JA biosynthesis enzymes upstream in the pathway, namely *LOX2/3* and *AOS*; however, these interactions have not been functionally validated (Pandey et al., 2008). Given that *AOC3* and *OPR3* mRNA abundance increased with JA, while their corresponding miRNAs decreased, such regulation would contribute to the known positive feedback loop of JA biosynthesis (Wasternack and Hause, 2013). Interestingly, although miR319 was identified in pokeweed, it was not JA-responsive. In tomato, JA treatment significantly reduced miR319 expression, which was associated with an increase in *TCP4* mRNA (Zhao et al., 2015). miR319 targets several class II TCPs, which function as transcriptional activators of *LOX2*

(Schommer et al., 2008). We hypothesize that the existence of pokeweed-specific, JA-responsive miRNAs reflects an alternative strategy to regulate JA biosynthesis. This hypothesis is further supported by the absence of miR319 binding sites in any TCP-encoding mRNA transcript in pokeweed. We also predicted that *ERF1B* is targeted by a miRNA; the encoded transcription factor is a key mediator of JA signal transduction that integrates ethylene-induced signals (Wasternack and Hause, 2013). To the best of our knowledge, this is the first report that *OPR3* mRNA is cleaved by a miRNA, and that *AOC3* and *ERF1B* are putative miRNA targets.

Apart from JA biosynthesis and signalling, we identified putative targets having more widespread roles in stress tolerance. These are summarized in **Table 11** and include several factors involved in salicylic or abscisic acid (ABA) signalling, with well-established roles in pathogen and environmental stress resistance, and others associated with classic growth hormones. Analysis of the expression dynamics of these targets provides insight into the role of miRNAs in mediating JA crosstalk with other hormone pathways. For example, we identified two miRNA/target pairs involving negative regulators of ABA signalling: Probable protein phosphatase 2C 73 and U-box domain-containing protein 19. Both genes had increased expression with JA, suggesting a dampening of the ABA pathway. This agrees with the known relationship between ABA and pathogen susceptibility, mediated by the transcription factor MYC2, which activates the ABA pathway and inhibits the JA pathway (Derksen et al., 2013). The ABA pathway contributes heavily to environmental stress tolerance (e.g. drought), so these miRNA/target interactions could modulate the trade-off between abiotic and biotic stress responses. A separate trade-off exists between stress response and growth, and JA inhibits developmental processes including seed germination and root/shoot growth (Wasternack and Feussner, 2017). We identified a miRNA/target pair comprising Auxin response factor 1, which is a repressor of auxin signalling

(Ellis et al., 2005). Expression of this gene increased with JA, suggesting inhibition of auxin-associated gene expression and related growth processes. The involvement of JA-responsive miRNAs in regulating Auxin response factor genes has been reported by others, indicating that these genes represent a conserved family of targets (Bozorov et al., 2012; Shen et al., 2017; Xu et al., 2016). Finally, we identified targets involved in redox activities: Glutathione S-transferase U9 and Isocitrate dehydrogenase, which may be relevant to the ability of pokeweed to hyperaccumulate heavy metals (Liu et al., 2010; Peng et al., 2008; Zhao et al., 2012). Notably, exogenous application of JA reduces heavy metal-associated oxidative stress (Singh and Shah, 2014). Taken together, our results suggest that miRNA-mediated regulation of the JA response contributes to widespread stress resistance activities in pokeweed. We chose to study pokeweed because it is a non-model plant known to resist biotic and abiotic pressure; however, the JA responsive miRNA/target pairs identified in this study may have agricultural relevance in terms of improving resistance to pathogen and environmental stresses in other plants.

### 3.6 Acknowledgement

The authors are grateful to Dr. Yutaka Amemiya, Manager, Genomics Core Facility, Next Generation Sequencing Lab, Sunnybrook Research Institute for small RNA sequencing and for his assistance and advice with sequence analysis.

# CHAPTER 4:

## De novo assembly of the pokeweed genome provides insight into pokeweed antiviral protein (PAP) gene expression

Published manuscript

**Neller, K. C. M.**, Diaz, C. A., Platts, A. E., and Hudak, K. A. (2019). *Front. Plant Sci.* doi: 10.3389/fpls.2019.01002

**Author contributions:** KN and KH conceived the study. KN performed bioinformatic analyses. CD performed cloning, gene model validations, measurement of gene expression from reporter constructs, and identification of promoter elements. AP provided advice on bioinformatic analyses and data interpretation. KN drafted the manuscript. KH edited the manuscript.

All Supplementary Material referenced in this manuscript (**Supplementary Figures 1-4**, **Supplementary Tables 1-2**, **Supplementary Data Sheets 1-6)** is available at the above doi.

## 4.1 Abstract

Ribosome-inactivating proteins are RNA glycosidases thought to function in defense against pathogens. These enzymes remove purine bases from RNAs, including rRNA; the latter activity decreases protein synthesis in vitro, which is hypothesized to limit pathogen proliferation by causing host cell death. Pokeweed antiviral protein (PAP) is a ribosome-inactivating protein synthesized by the American pokeweed plant (*Phytolacca americana*). PAP inhibits virus infection when expressed in crop plants, yet little is known about the function of PAP in pokeweed due to a lack of genomic tools for this non-model species. In this work, we de novo assembled the pokeweed genome and annotated protein-coding genes. Sequencing comprised paired-end reads from a short-insert library of 83X coverage, and our draft assembly (N50 = 42.5 Kb) accounted for 74% of the measured pokeweed genome size of 1.3 Gb. We obtained 29,773 genes, 73% of which contained known protein domains, and identified several PAP isoforms. Within the gene models of each PAP isoform, a long 5' UTR intron was discovered, which was validated by RT-PCR and sequencing. Presence of the intron stimulated reporter gene expression in tobacco. To gain further understanding of PAP regulation, we complemented this genomic resource with expression profiles of pokeweed plants subjected to stress treatments (jasmonic acid, salicylic acid, polyethylene glycol, and wounding). Cluster analysis of the top differentially expressed genes indicated that some PAP isoforms shared expression patterns with genes involved in terpenoid biosynthesis, jasmonic acid-mediated signalling, and metabolism of amino acids and carbohydrates. The newly sequenced promoters of all PAP isoforms contained cis-regulatory elements associated with diverse biotic and abiotic stresses. These elements mediated response to jasmonic acid in tobacco, based on reporter constructs containing promoter truncations of PAP-I,

the most abundant isoform. Taken together, this first genomic resource for the Phytolaccaceae plant family provides new insight into the regulation and function of PAP in pokeweed.

## 4.2 Introduction

American pokeweed, *Phytolacca americana*, belongs to the Phytolaccaceae family of flowering plants, which comprises 65 species of herbs, shrubs, and trees. *P. americana* (pokeweed) is the most well-studied *Phytolacca* species due to its broad agricultural and medical applications. Pokeweed synthesizes pokeweed antiviral protein (PAP), an N-glycosidase and ribosome-inactivating protein (RIP) that depurinates the conserved α-sarcin loop of large rRNAs (Endo et al., 1988). PAP exhibits antiviral activity against diverse plant and animal viruses. Specifically, depurination of rRNA inactivates ribosomes in infected cells, thereby inhibiting host and viral protein synthesis (Lodge et al., 1993; Rajamohan et al., 1999). PAP also depurinates the genomes of some RNA viruses, interfering with multiple stages of the viral life cycle (He et al., 2008; Karran and Hudak, 2008; Mansouri et al., 2009). Transgenic plants expressing PAP acquire novel antiviral and antifungal activities, making the gene an attractive candidate for use in agricultural engineering (Dai et al., 2003; Wang et al., 1998; Zoubenko et al., 1997, 2000). Pokeweed also shows potential in phytoremediation as a heavy metal hyperaccumulator, thriving in contaminated soil that is otherwise toxic to most plants (Liu et al., 2010; Peng et al., 2008; Zhao et al., 2011). Although this non-model plant displays resistance to diverse biotic and abiotic stresses, a wealth of information remains unknown since its genome has not been sequenced.

Several isoforms of PAP have been reported, exhibiting different temporal (PAP-I, PAP-II, PAP-III) or tissue-specific (PAP-R, PAP-S, S1, S2, PAP-α) expression patterns, or identified during cell culture (PAP-H, PAP-C) (Barbieri et al., 1982, 1989; Bolognesi et al., 1990; Irvin, 1975; Irvin et al., 1980; Kataoka et al., 1992; Park et al., 2002; Rajamohan et al., 1999). Consistent with a hypothesized role in pathogen defense, we showed previously through transcriptomic analysis that expression of several PAP isoforms is up-regulated by jasmonic acid (JA) (Neller et

al., 2016). JA is a plant hormone that mediates resistance to insect herbivores, which are viral vectors, and necrotrophic pathogens. Others have reported an induction of RIP expression in various plants upon treatment with phytohormones (JA, salicylic acid (SA), abscisic acid (ABA)) or associated stresses, including insect feeding, pathogen infection, cold, heat, drought, salinity, and mechanical wounding (Iglesias et al., 2005; Jiang et al., 2008; Qin et al., 2009; Reinbothe et al., 1994; Song et al., 2000; Tartarini et al., 2010). Therefore, it is well-established that RIPs are induced by various stresses; however, it is not clear how RIP isoform expression is controlled, or how RIPs are integrated within stress-response pathways.

Pokeweed is tetraploid with a chromosome count of 2n = 36 and 1C-value of 1.48 pg (~1.5 Gb) (Bennett, 2000; Rice et al., 2015). Although there are no reference genomes available for the Phytolaccaceae family, genomes of some members from the order Caryophylalles (which includes pokeweed) have been sequenced. Genome assembly and annotation has been performed for *Beta vulgaris* (sugar beet), *Spinacia oleracea* (spinach), *Chenopodium quinoa* (quinoa), and *Amaranthus hypochondriacus* (amaranth) (Clouse et al., 2016; Dohm et al., 2014; Jarvis et al., 2017; Xu et al., 2017). This genomic information is a useful reference for pokeweed, especially since RIPs are prevalent among the Caryophyllales (Di Maro et al., 2014).

Here, we present the first de novo draft genome assembly of pokeweed and an annotation of its protein-coding genes. Using this resource, we investigated the presence and gene organization of PAP isoforms. A novel feature was discovered in the 5' untranslated region (UTR): a long intron that affects gene expression. Integration of RNA-Seq data from pokeweed stress treatments enabled the identification of co-expressed genes and provided insight into defense responses in the plant. Finally, differences in cis-regulatory elements (CREs) in the promoters of PAP isoforms, combined with their unique expression profiles, suggest that the isoforms have

distinct roles in pokeweed. Our study provides a workflow that integrates genome assembly, annotation, and differential expression analysis for a non-model plant. These resources will facilitate the study of pokeweed, to understand its ability to survive environmental stress.

## 4.3 Material and methods

### Genomic DNA sequencing and genome assembly

A summary of our study is provided in **Figure 16**. High-quality genomic DNA was isolated from a single pokeweed plant using a CTAB-based extraction method (Healey et al., 2014) and sent to McGill University and Genome Quebec Innovation Centre (Montreal, QB, Canada) for sequencing. Shotgun sequencing was performed on one lane of an Illumina HiSeq 2500 instrument in Rapid Run mode. Paired-end reads of 250 bp were obtained from genomic DNA fragments having an average size of 400 bp; this strategy was chosen to conform with input recommendations of the downstream assembly software. Following sequencing, adapters and low-quality bases ($Q < 30$, averaged over four bases) were removed with Trimmomatic (v. 0.36; Bolger et al., 2014). The pokeweed genome was assembled with Discovar De Novo using default parameters (v. 52488; Love et al., 2016) on the Sharcnet high-performance computing cluster (https://www.sharcnet.ca). Raw genomic DNA sequencing reads are available at the SRA under project # PRJNA544344. The genome completion score was measured with BUSCO (v. 3; Waterhouse et al., 2018).

### Determination of genome size

DNA contents of somatic nuclei (pg/2C) were measured using flow cytometry. For each sample, approximately 1 $cm^2$ of fresh pokeweed leaf and 2 $cm^2$ of fresh *Sorghum bicolor* Pioneer

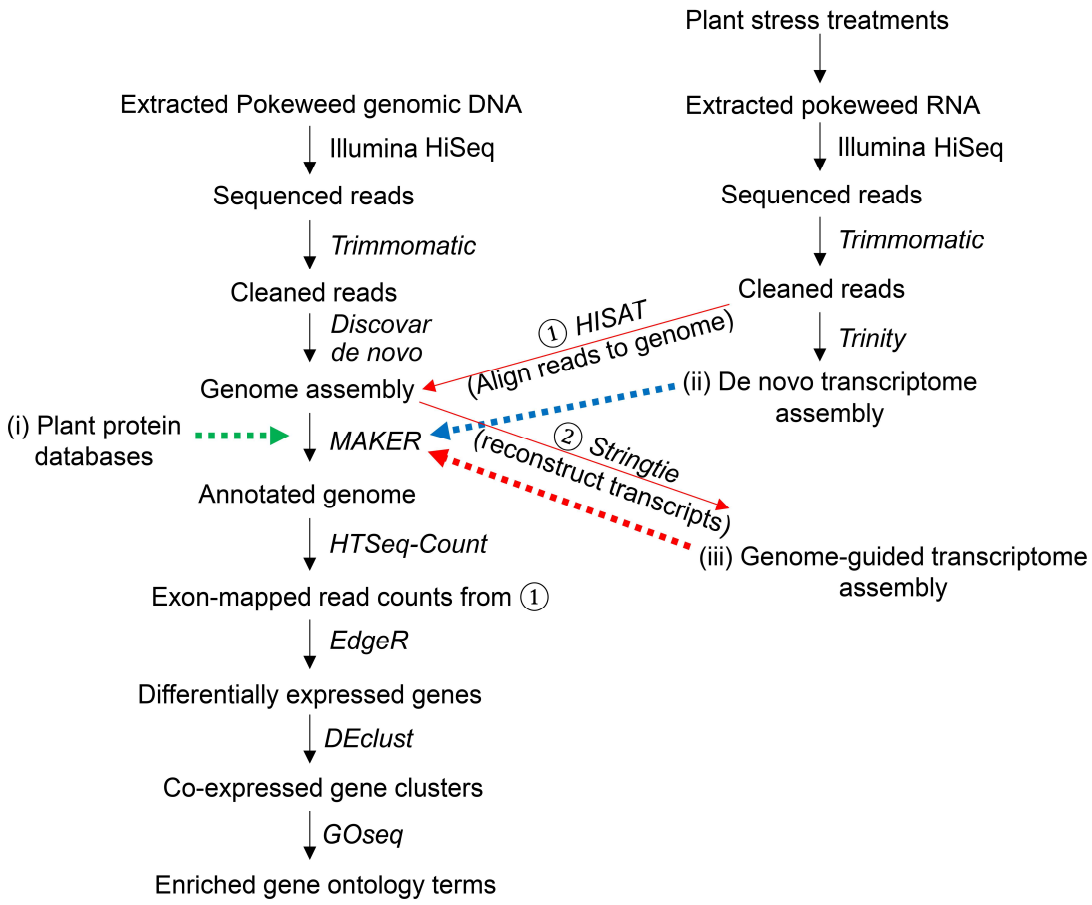



**Figure 16. Summary of genome assembly, annotation, and differential gene expression analysis**. The assembled pokeweed genome was annotated using three sources of information provided to MAKER, indicated with dashed arrows: (i) published plant protein databases, (ii) the pokeweed de novo transcriptome assembly, (iii) the pokeweed genome-guided transcriptome assembly, derived from aligning unassembled ('cleaned') reads to the genome (①) and joining exonic regions into transcripts (②). Reads from (①) that aligned to exonic regions of a gene (defined by the annotated gene models) were summed. Counts were used as input for differential gene expression analysis. The top differentially expressed genes were clustered, and functional enrichment testing was performed on each cluster.

8695 (1.74 pg/2C; Johnston et al., 1999) were chopped with a razor blade in 0.7 mL of ice-cold LB01 buffer (Dolezel et al., 1989) containing 100 µg/mL propidium iodide and 50 µg/mL RNAse A. Samples were stained at room temperature for 20 to 25 minutes, then tested at low speed using a BD FACSCalibur flow cytometer (BD Biosciences, San José, USA). The FL2 detector (585/42nm) was used to measure fluorescence, with integrated fluorescence as the parameter of interest. 2C DNA contents were calculated based on the relative fluorescence of the pokeweed and *Sorghum* G0/G1 nuclei and the known DNA content of *Sorghum*. Means, coefficients of variation, and nuclei numbers for the nuclei fluorescence peaks were measured using flowPloidy (v. 1.7.0; Smith et al., 2018). Nine pokeweed plants were tested in total: four as individuals and five as bulk samples of two or three plants.

**Stress treatments, mRNA sequencing, and transcriptome assembly**

Total RNA was extracted from four-leaf pokeweed plants subjected to the following treatments: sprayed with 5 mM JA or SA (solubilized in 0.5% ethanol), watered every three days for a seven-day period with 10% polyethylene glycol (PEG), or wounded with forceps (WND). Plants sprayed with 0.5% ethanol (ET) or watered normally (WT) served as controls for JA/SA and PEG/WND, respectively. Leaf tissue was flash-frozen in liquid nitrogen 24h following treatment for JA, SA, ET, and WND samples, and three days after the final treatment for PEG and WT samples. RNA-Seq libraries were constructed with the TruSeq Stranded mRNA Library Preparation Kit (RS-122-2101, Illumina). For each condition, four biological replicate libraries were prepared from equal amounts of RNA pooled from three independent plants (i.e. 24 libraries derived from 72 total plants). Strand-specific, paired-end reads of 125 bp were sequenced on two lanes of an Illumina HiSeq 2500 instrument by The Centre for Applied Genomics (The Hospital

for Sick Children, Toronto, ON, Canada). Raw mRNA sequencing reads are available at the SRA under project # PRJNA309999.

In addition to the data generated in the present study, mRNA transcriptome assembly incorporated two publicly available *P. americana* RNA-Seq datasets from the SRA: Accessions SRX2774676 and ERX2099309, which were derived from whole plants (root, stem, leaf, and flower tissue), as well as high-coverage RNA-Seq datasets (n=3 biological replicates) from our previous study of the pokeweed leaf mRNA transcriptome (Neller et al., 2016). All datasets (32 in total) were processed with Trimmomatic as described above. Two transcriptome assembly strategies were employed. For de novo assembly, reads were combined into a single reference transcriptome using Trinity (v. 2.5.1; Grabherr et al., 2011; Haas et al., 2013). For genome-guided assembly, RNA-Seq libraries were independently aligned to the assembled genome with HISAT (v. 2.1.0; Kim et al., 2015) and transcript reconstruction was performed with Stringtie (v. 1.3.4b; Pertea et al., 2015). The independent assemblies were combined into a single, non-redundant reference transcriptome using the Stringtie 'merge' function.

**Genome annotation**

Only contigs of minimum length 10 Kb were annotated, as these were most likely to contain full-length protein-coding genes (Campbell et al., 2014a). Prior to annotation, a pokeweed-specific repeat library was prepared with RepeatModeler (v. 1.0.9; http://www.repeatmasker.org/RepeatModeler/). Repeat-masking and gene prediction were performed using the MAKER pipeline (v. 2.31.8; Campbell et al., 2014a, 2014b). Both pokeweed-specific and simple repeats were masked with RepeatMasker (v. 4.0.7;

http://www.repeatmasker.org/). RepeatRunner was used to mask divergent protein-coding portions of retro-elements and retro-viruses not identified by RepeatMasker (Smith et al., 2007).

Gene prediction followed the protocol outlined in Campbell et al., 2014a. The Trinity (de novo) and Stringtie (genome-guided) transcriptomes described above were provided as transcript evidence. The proteomes of published Caryophyllales species (sugar beet, spinach, quinoa, amaranth) were used as homologous protein evidence (Clouse et al., 2016; Dohm et al., 2014; Jarvis et al., 2017; Xu et al., 2017). The quinoa and amaranth proteomes were obtained from Phytozome (https://phytozome.jgi.doe.gov; v. 1.0 and v. 2.1, respectively), spinach from SpinachBase (v. 1; http://www.spinachbase.org), and sugar beet from The *Beta vulgaris* Resource (Refbeet-1.1; http://bvseq.boku.ac.at). Protein evidence was further supplemented with the plant subset of the SwissProt database (obtained October 2017; Pundir et al., 2017). Three rounds of iterative gene prediction were performed with MAKER. In Round 1, predictions were inferred directly from transcript and protein evidence (est2genome, protein2genome = 1). Top-scoring gene models from Round 1 (AED $\leq$ 0.25, amino acids $\geq$ 50) were used to train the ab initio gene predictors SNAP (Korf, 2004) and AUGUSTUS (v. 3.2.3; Stanke et al., 2004). In Round 2, MAKER was re-run with ab initio gene predictors turned on, and top-scoring models were used to re-train predictors as per above. A final Round 3 was run with the twice-trained gene predictors.

Following gene prediction, an Interproscan (v. 5.28-67.0; Zdobnov and Apweiler, 2001) search was performed to identify protein-coding (Pfam) domains and obtain associated gene ontology (GO) terms. GO annotations were supplemented with non-redundant GO terms from the SwissProt database for the best hit as per BLAST-P analysis. Each round of genome annotation was inspected visually with JBrowse (v. 1.12.3 Buels et al., 2016). In Rounds 2 and 3, we observed spurious fusion of protein-coding gene models with putative pseudogenes; this could not be

resolved despite several attempts of re-training and post-annotation processing. Therefore, the final gene set was conservatively chosen to comprise only the evidence-based annotations from Round 1. The completion score of the annotated gene set was measured with BUSCO as described above, as well as relative to CoreGFs of the 'green plants' subset from PLAZA v. 2.5 (Veeckman et al., 2016).

**Validation of PAP isoform gene models**

cDNAs were generated by reverse transcribing 0.5 µg of total pokeweed RNA with SuperScript III reverse transcriptase (25 units; Thermo Fisher) and isoform-specific primers. To validate PAP isoform gene models at both the genomic and mRNA levels, PCR was performed with isoform-specific forward and reverse primers using either pokeweed gDNA or cDNA as the starting template. All PCR amplifications were conducted using Q5 High-Fidelity DNA polymerase (1 unit; New England Biolabs) and 0.5 µM of each primer. PCR products were gel-purified and used as templates for a second round of PCR, this time with forward and reverse primers that contained additional sequences for cloning. Amplicons from the second round of PCR were gel-purified and cloned into the multiple cloning site of pHSG299 vector using the one-step SLIC (sequence- and ligation-independent cloning) method (Jeong et al., 2012). All constructs were sequenced and compared with computationally-derived gene and mRNA models. Primers used in this study are listed in **Supplementary Table 1**.

**Orthogroup analysis**

The longest isoform per gene was obtained for each Caryophyllales proteome noted above, and orthogroup assignment was performed with OrthoFinder (v. 2.2.6; Emms and Kelly, 2015).

Functional enrichment analysis of genes from pokeweed-specific orthogroups was conducted with the GOseq package (Young et al., 2010) in the statistical program R (R Foundation for Statistical Computing, 2016). Pokeweed-specific genes were tested for enriched GO terms relative to all pokeweed genes.

**Differential gene expression analysis**

HTSeq (v. 0.8.0; Anders et al., 2015) was used to sum exon-level counts per gene for each independent RNA-Seq library aligned to the pokeweed genome. Gene counts were provided as input for differential expression testing in the Bioconductor package EdgeR (v. 3.7; Robinson et al., 2010). All pairwise tests were performed with four biological replicates per treatment, and genes with FDR < 0.05 in at least one comparison were considered differentially expressed. Treatment-responsive genes were identified based on the relevant tests: ET vs. JA or SA, and WT vs. PEG or WND. Top differentially expressed genes (DEGs) were defined as those with FDR < 0.001 and FC > 4 in at least one pairwise comparison. These genes were clustered with DEclust (v. 1.0.1; Aoto et al., 2017) based on a multi-conditional expression profile of EdgeR results and normalized abundance in transcripts per million (TPM). GO enrichment analysis was performed on the genes in each cluster as described above.

**Quantitative RT-PCR (qRT-PCR) validation of PAP isoform mRNA levels**

cDNAs were generated by reverse transcribing 0.5 µg of total pokeweed RNA with SuperScript III reverse transcriptase (25 units; Thermo Fisher) and gene-specific primers (Supplementary Table 1) according to the manufacturer's instructions. Elongation factor-1-gamma (EF1G) and the cell wall protein BIIDXI served as internal controls as these transcripts were stably

expressed under our stress treatments according to our RNA-Seq differential expression analysis. Following cDNA synthesis, 5 µL of the reverse transcription (RT) reaction product was combined with 0.3 µM forward primer, 0.3 µM reverse primer, and 33 µL of 2X SYBR Green qPCR Master Mix (Bimake), to a final volume of 66 µL. Each reaction was split into three technical replicates. qRT-PCRs were conducted in a QIAGEN RotorGene Q thermocycler with the following settings: hold at 50°C for 20 seconds, initial denaturation and hot-start DNA polymerase activation for 10 minutes, followed by 40 cycles alternating between denaturation (95°C; 15 seconds) and combined annealing/extension (68°C, 45 seconds). mRNA levels were quantified using the ΔΔCt method (Livak and Schmittgen, 2001). A melting curve analysis was performed to confirm the presence of a single PCR product after each reaction. At least three biological replicates per treatment were conducted for each transcript.

**Generation of promoter-GUS reporter constructs**

The 1262 bp region upstream of the PAP-I transcription start site (TSS) was considered the proximal PAP-I promoter. This ~1.3 Kb promoter, along with the 5' UTR and 1.6 Kb intron, were PCR-amplified from pokeweed genomic DNA (500 ng) with the primers PAP-I-prom-SLIC-FOR and PAP-I-prom-SLIC-REV (0.5 µM each). Amplicons were gel-purified and cloned by one-step SLIC (Jeong et al., 2012) into the multiple cloning site of pHSG299. All PCR amplifications were conducted using Q5 High-Fidelity DNA polymerase (1 unit; New England Biolabs) using the manufacturer's instructions. The pHSG299 plasmid containing the PAP-I promoter and intron then served as the PCR template for all downstream PCRs of PAP-I promoter fragments. PAP-I promoter fragments containing the 1.6 Kb intron (1262-int and 102-int) were generated through

PCR by pairing the same reverse primer (5-UTR-SLIC-REV) with different forward primers (P1-FOR or P7-FOR). Primer sequences are listed in **Supplementary Table 1**.

To produce intronless versions of promoter constructs, a reverse primer (5-UTR-no-int-REV) was designed to connect the two portions of the PAP-I 5' UTR that were originally interrupted by the intron. This reverse primer was paired with different forward primers (P1-FOR to P7-FOR) to produce 5' promoter truncations (1262, 1124, 711, 584, 432, 296, and 102). PCR products were gel-purified and used as templates for a second round of PCR to attach sequences needed for SLIC cloning.

Amplified fragments were cloned in place of the CaMV 35S promoter in pCambia 0305.2 using one-step SLIC (Jeong et al., 2012) and transformed into *Escherichia coli* DH5α. For constructs used to investigate the effect of the PAP-I leader intron on gene expression, the castor bean catalase intron in the GUS reporter gene was removed to determine the influence of the PAP-I intron alone. Minus catalase intron constructs were made by excluding the intron through PCR (Primer pairs: pCambia-1-no-cat-FOR and pCambia-1-no-cat-REV; pCambia-2-no-cat-FOR and pCambia-2-no-cat-REV) and reassembling the two fragments through Gibson assembly (Cat# E2611S; New England Biolabs). Positive *E. coli* transformants were screened through colony PCR, and all reporter gene constructs were confirmed by sequencing.

Constructs were transformed into *Rhizobium radiobacter* (syn. Agrobacterium) AGL1 strain by electroporation as previously described (Wise et al., 2006). Electrocompetent cells (20 µL) mixed with plasmid DNA (50 ng) were electroporated (2.5 kV, 25 µF capacitance, and 400 Ω resistance) and allowed to recover in 2 mL of non-selective YEP medium for 2h. After recovery, 100 µL of cells were plated on selective YEP agar (50 µg/mL carbenicillin, 50 µg/mL kanamycin) and incubated at 28°C to allow colonies to form.

**Measurement of reporter activity (GUS histochemical and fluorometric assay)**

Agrobacterium cultures harbouring promoter-reporter gene constructs were used to agroinfiltrate leaves of four-leaf stage *Nicotiana tabacum* plants as previously described (Zhao et al., 2017). Cultures were grown in selective YEP liquid medium (50 µg/mL carbenicillin, 50 µg/mL kanamycin) until late log phase ($OD_{600}$ = 0.7 – 1.0). Cells were then pelleted by centrifugation, washed in agroinfiltration solution (10 mM MES-KOH, pH 5.6, 10 mM $MgCl_2$, 200 µM acetosyringone), and resuspended in agroinfiltration solution to a final $OD_{600}$ of 0.5. Agrobacterium cells were injected into the abaxial surface of leaves using a needleless syringe. For the JA experiment, leaves were treated with either 0 mM JA (0.5% ethanol; mock) or 5 mM JA 24h after agroinfiltration. Leaf discs from inoculated plants were harvested 72h post agroinfiltration and either used directly for histochemical assays or stored in liquid nitrogen until processing for the fluorometric assay.

GUS histochemical assays were performed according to Jefferson et al. (1987) with some modifications. Fresh leaf discs (0.5 cm diameter) from inoculated plants (minimum of three plants per construct) were vacuum-infiltrated with 5-bromo-4-chloro-3-indolyl-b-D-glucuronic acid, cyclohexylammonium salt (X-Gluc) solution, and incubated overnight at 37°C. After GUS staining, leaf discs were cleared of chlorophyll by washing in increasing concentrations of ethanol (70%-100%) for 48h.

GUS fluorometric assays were performed in black, clear-bottom 96-well plates according to Côté and Rutledge (2003), with some modifications. Frozen leaf discs (1 cm diameter; 2 discs per sample) from inoculated plants (minimum of four plants per construct) were combined with 200 mg of glass beads (1.0 mm, BioSpec) and 300 µL of GUS extraction buffer (50 mM phosphate buffer, pH 7.0, 10 mM DTT, 10 mM EDTA, 0.1% SDS, 0.1% Triton X-100, 10 mM β-

mercaptoethanol), and homogenized using a 3110BX MiniBeadBeater (BioSpec) for 120 s at speed 48. Tissue debris was removed by centrifugation at 4ºC and cleared plant extracts (10 µL) were mixed with 720 µL of 0.1 mM 4-methylumbelliferyl-β-D-glucuronide hydrate (4-MUG) and incubated at 37ºC. Beginning at 0 minutes, a 10-µL aliquot was taken from each reaction every 15 minutes for a total of 60 minutes, and pipetted into a well containing 180 µL of stop buffer (0.2 M $Na_2CO_3$). Three technical replicates per time point were taken for each sample. Fluorescence values were measured at room temperature using a Synergy H4 Hybrid microplate reader (excitation: 365 nm; emission: 455 nm) and compared to a previously determined 4-methylumbelliferone (4-MU) standard curve. GUS activity was calculated from the linear slope of the fluorescence readings and normalized to the total protein concentration, which was determined using a BCA Reducing Agent Compatible Protein Assay Kit (G-Biosciences). Comparisons between mock-treated and JA-treated samples ($p < 0.01$) were performed for each promoter construct using two-tailed t-tests.

**Identification of putative PAP promoter CREs and JA-responsive transcription factors**

As with PAP-I, the 1.3 Kb sequence upstream of the TSS was considered the proximal promoter for each PAP isoform. Promoter sequence identity analysis was performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters. Putative plant-specific cis-regulatory elements (CREs) were identified using PLACE (http://www.dna.affrc.go.jp/htdocs/PLACE/) and PlantPan 2.0 (http://plantpan2.itps.ncku.edu.tw/) web interfaces (Chow et al., 2016; Higo et al., 1999). To reduce false positives, only sequences with ≥ 90% identity to the published motifs were included. Putative transcription factors (TFs) associated with CREs were identified in the pokeweed genome

based on annotated Pfam domains, and their differential expression results were assessed for JA-responsiveness (FDR < 0.05).

## 4.4 Results

### Assessment of the pokeweed genome assembly

Statistics of the pokeweed genome assembly are shown in **Table 12**. Based on 1 Kb+ contigs, the assembly is 0.93 Gb in size and has a read coverage of 83X. The contig and scaffold N50 values are 35.2 Kb and 42.5 Kb, respectively, where a scaffold represents the single highest coverage path through each line of the genome assembly graph. The scaffold assembly was used for all downstream analysis. The assembly had a BUSCO genome completion score of 84.3%; of a possible 1440 plant BUSCOs, 1214 were identified as complete (1142 single-copy and 72 duplicated), 82 were fragmented, and 144 were missing. For an additional metric of assembly completion, we determined the expected size of the pokeweed genome using flow cytometry (**Supplementary Figure 1**). Based on triplicate samples of four individual plants and two non-replicate bulk samples, the pokeweed plants used in this study had a genomic content (mean ± SEM) of 2.57 pg/2C ± 0.0019. Variability was low, with all 14 samples measuring either 2.57 or 2.58 pg/2C. Pokeweed and *Sorghum* nuclei fluorescence peaks had coefficients of variation < 3.3% and included data from at least 1,000 nuclei. Given the conversion that 1 pg DNA = 980 Mb, the haploid genome size of pokeweed is estimated to be 1.26 Gb. Therefore, the present assembly represents 74% of the expected genome size.

**Table 12. Statistics of the pokeweed genome assembly.**

| | |
|---|---|
| Number of input reads | 330,991,096 |
| Number of edges | 1,829,078 |
| Mean edge length (bases) | 1,111 |
| Contig line N50 (bases) | 35,208 |
| Number of scaffolds | 847,766 |
| Scaffold line N50 (bases) | 42,514 |
| Total bases in 1 Kb+ scaffolds | 933,313,610 |
| Total bases in 10 Kb+ scaffolds | 821,040,380 |
| Number of 10 Kb+ scaffolds | 22,292 |
| Number of 1Kb+ scaffolds | 70,834 |
| Longest scaffold length (bases) | 366,732 |
| Shortest scaffold length (bases) | 200 |
| Estimated read coverage | 83X |
| BUSCO completion score | 84% |

**Annotation of the pokeweed genome**

Scaffold contigs of minimum length 10 Kb were annotated with MAKER, as these were most likely to contain protein-coding genes (Campbell et al., 2014a). Accordingly, the mapping rate of pokeweed RNA-Seq reads aligned to these 10 Kb+ contigs averaged 93% over all samples. The annotation file and sequences of annotated transcripts and proteins are provided in **Supplementary Data Sheets 1-3**. A summary of the annotation is shown in **Table 13**. From 22,292 contigs, we obtained 29,773 genes (mean length = 5,072 bp) and 56,538 mRNA transcripts (mean length = 1,720 bp). Importantly, 73% of genes contained a Pfam domain, indicating that the majority are protein-coding, and 99% of genes had an AED score < 0.5 (**Supplementary Figure 2**), demonstrating excellent correspondence between evidence and gene models. The latter result is consistent with our evidence-only genome annotation; that is, gene models were based solely on the provided transcript and protein information rather than ab initio prediction.

Although we had attempted to include iterative training and ab initio prediction, doing so resulted in fused gene models that could not be validated by RT-PCR and PCR. Specifically, following two iterative training rounds we obtained an increase of only 72 genes (0.2%) but a substantial (30%) increase in mean gene length. Visual inspection of the annotations suggested that protein domain-containing pseudogenes had become fused with bona fide genes. This issue of spurious gene fusions upon ab initio incorporation persisted despite varying the gene-finding parameters, using MAKER post-processing tools, and applying the AUGUSTUS species model from the well-annotated sugar beet genome. Surprisingly, we observed that ab initio incorporation still resulted in high AED scores: after two iterative training rounds, 95% of genes had an AED < 0.5 (**Supplementary Figure 2**); this reinforces the importance of visibly inspecting gene models and associated evidence. Given that ab initio gene prediction resulted in only a small number of

**Table 13. Statistics of the annotated pokeweed genome.**

| | |
|---|---|
| Number of genes | 29,773 |
| Number of mRNAs | 56,538 |
| Gene length | Mean: 5,072<br>Median: 3,452 |
| Transcript length | Mean: 1,720<br>Median: 1,513 |
| Exon length | Mean: 278<br>Median: 152 |
| Intron length | Mean: 898<br>Median: 389 |
| CDS length | Mean: 1,121<br>Median: 921 |
| 5’ UTR | Mean: 320<br>Median: 241 |
| 3’ UTR | Mean: 457<br>Median: 377 |
| Exons per mRNA | Mean: 5<br>Median: 6 |
| % of genes with Pfam domain | 73% |
| % of genes with SwissProt hits | 64% |

novel genes at the expense of annotation accuracy, we chose to conduct all downstream analyses with our evidence-based gene set.

**Evaluation of the annotated gene set**

The BUSCO completion score of the annotated pokeweed gene set was 75.8%. Since the genome assembly received a completion score of 84.3% (above), 90% of the BUSCOs identified in the genome were annotated. We also measured completion of the gene set relative to CoreGFs of the 'green plants' subset from PLAZA. The CoreGF reference set has a lower threshold of species conservation than BUSCO and is not limited to single-copy genes (Veeckman et al., 2016). Based on CoreGF analysis, the gene set completion score was 97.6%. Therefore, in the case of a non-model species, a low BUSCO score may be more reflective of high evolutionary divergence from reference species than incomplete genome assembly or annotation.

To assess how the pokeweed gene set compared to that of previously annotated Caryophyllales species, we used OrthoFinder to identify orthologous gene families (orthogroups) among pokeweed, sugar beet, quinoa, spinach, and amaranth. The species-specific distribution of orthogroups is shown in **Figure 17** and **Supplementary Table 2**. Pokeweed genes were distributed into 14,785 orthogroups comprised of 22,901 genes. Eighteen orthogroups (116 genes) were pokeweed-specific, and these genes were enriched in the GO terms 'far-red light signaling pathway' (FDR = 0.0041) and 'negative regulation of defense response' (FDR = 0.058). The respective genes were annotated as isoforms of the TF FAR1-Related Sequence (FRS) (PHYAM_025199, PHYAM_007976, PHYAM_022646, PHYAM_016237, PHYAM_019944, PHYAM_019825) and the F-box protein Constitutive Expresser of PR genes 30 (CPR30) (PHYAM_017382, PHYAM_000825, PHYAM_000824, PHYAM_016094). For pokeweed, 77%

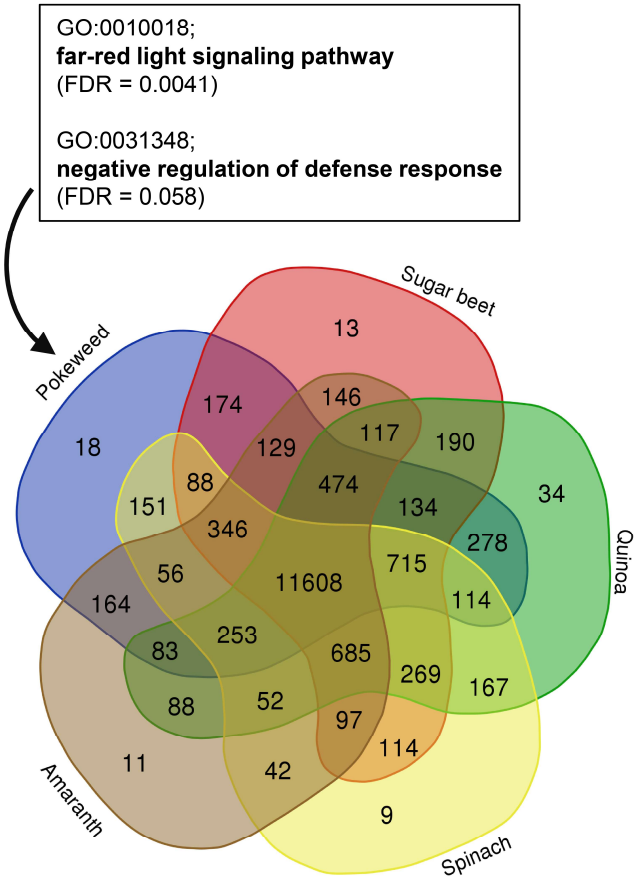



**Figure 17. Identification of pokeweed-specific genes by orthogroup analysis of Caryophyllales species.** Orthogroup assignment was performed with OrthoFinder, using the longest representative protein per gene for each species. The species distribution of orthogroups is shown. Enriched GO terms from pokeweed-specific genes (n=116) are indicated in the box.

of genes were assigned to orthogroups and 0.4% of genes were species-specific. This is in line with other Caryophyllales species, for which orthogroup-assigned genes ranged from 72% (sugar beet) to 82% (amaranth) and species-specific genes ranged from 0.1% (spinach) to 0.4% (quinoa). Taken together, these results demonstrate that the pokeweed gene set is consistent with that of well-annotated Caryophyllales species. Additionally, we have identified a subset of genes that may contribute to distinct biological relevance in pokeweed.

**Identification of RIP genes in pokeweed**

Following genome annotation, we identified genes containing a RIP domain by performing an Interproscan search of pokeweed protein sequences against the Pfam database. This analysis revealed 10 RIP domain-containing genes, summarized in **Table 14**. PAP isoform annotation was made based on a BLAST-P search against the SwissProt database, retaining the best-scoring hit per protein. Three RIP domain-containing genes had 100% identity and coverage with SwissProt sequences of PAP-I, PAP-α, and PAP-S, respectively. PAP-II was also present with 99% identity and 100% coverage. One gene had 77% identity and 100% coverage with PAP-I from SwissProt, but BLAST-N against pokeweed nucleotide sequences from GenBank showed 99% identity and coverage with the partial genomic clone of PAP-S2 (Accession # AB071855.1). Furthermore, a gene annotated as PAP-S above, had 99% identity and 100% coverage with the partial genomic clone of PAP-S1 (Accession # AB071854.1). Therefore, two PAP-S isoforms exist, namely PAP-S1 and PAP-S2, which agrees with a previous finding (Honjo et al., 2002) and clarifies the single PAP-S notation in SwissProt. We also identified a gene encoding a transcript that we previously reported (c18776_g1_i1) as a potential new PAP isoform (Neller et al, 2016). The gene contains a

**Table 14. Annotation of ribosome-inactivating protein genes in pokeweed.**

| Gene ID | Scaffold ID | BLAST-P Top Hit | Annotated isoform |
|---|---|---|---|
| PHYAM_020596 | line_15290<br>Pos: 5980-9049 | RIP1_PHYAM<br>ID=100%; Cov=100% | PAP-I |
| PHYAM_028184 | line_34803<br>Pos: 2685-10711 | RIP2_PHYAM<br>ID=99%; Cov=100% | PAP-II |
| PHYAM_021314 | line_16387<br>Pos: 24092-28770 | RIPS_PHYAM<br>ID=100%; Cov=100% | PAP-S1 |
| PHYAM_010467 | line_4604<br>Pos: 41936-48641 | RIP1_PHYAM<br>ID=77%; Cov=100% | PAP-S2 |
| PHYAM_022058 | line_17664<br>Pos: 4308-7191 | RIPA_PHYAM<br>ID=100%; Cov=100% | PAP-Alpha |
| PHYAM_012451 | line_6055<br>Pos: 38961-39890 | RIPA_PHYAM<br>ID=38%; Cov=96% | Novel isoform |
| PHYAM_008178 | line_3146<br>Pos: 67972-68871 | RIP2_PHYDI<br>ID=96%; Cov=100% | Non-transcribed pseudogene |
| PHYAM_008179 | line_3146<br>Pos: 74851-75880 | RIP1_PHYAM<br>ID=74%; Cov=100% | Non-transcribed pseudogene |
| PHYAM_010465 | line_4604<br>Pos: 18489-18926 | RIP1_PHYAM<br>ID=79%; Cov=46% | Transcribed pseudogene with in-frame stop codons |
| PHYAM_010468 | line_4604<br>Pos: 65729-65995 | RIPL2_PHYDI<br>ID=80%; Cov=32% | Transcribed pseudogene with in-frame stop codons |

RIP domain but only has 38% sequence identity and 96% coverage with its best hit, PAP-α, which supports this gene as a novel PAP isoform.

Our analysis also led to the identification of four RIP domain-containing genes that are likely PAP pseudogenes (**Table 14**). Two transcribed pseudogenes were located on the same contig as PAP-S2. Their gene models were truncated relative to the associated transcript evidence, resulting in 46% and 32% coverage with the respective hits by BLAST-P. Upon closer inspection, in-frame stop codons were observed in all reading frames for both genes; this explained the shortened models, which arose from annotating the longest open reading frame per gene. Two other probable pseudogenes were identified, but their transcripts were absent in either one or both transcriptome assemblies, suggesting that the genes are transcriptionally silent. To the best of our knowledge, this is the first report of potential pseudogenes of PAP.

**Presence of a long leader intron in gene models of PAP isoforms**

Gene models of protein-coding PAP isoforms are shown in **Figure 18A**. With exception of PAP-II, the coding sequence of all isoforms was annotated as a single exon of ~900 bp. In contrast, the coding sequence of PAP-II comprised two exons, separated by an intron of 736 bp, which agrees with a previous report (Poyet, 1997). Interestingly, the gene models of all isoforms revealed a long intron within the 5' UTR, ranging from 1.5 Kb (PAP-α) to 5.7 Kb (PAP-II). Based on the distribution of intron lengths in pokeweed, the PAP leader introns were longer than 83% (PAP-α) to 94% (PAP-II) of all introns. Several isoforms (PAP-II, PAP-S1, PAP-S2) were predicted to have multiple gene models that differed only in their 5' UTRs, which suggested the use of alternate promoters in some cases ($PAP\text{-}II_A$/$PAP\text{-}II_B$; $PAP\text{-}S2_A$/$PAP\text{-}S2_B$). In addition to

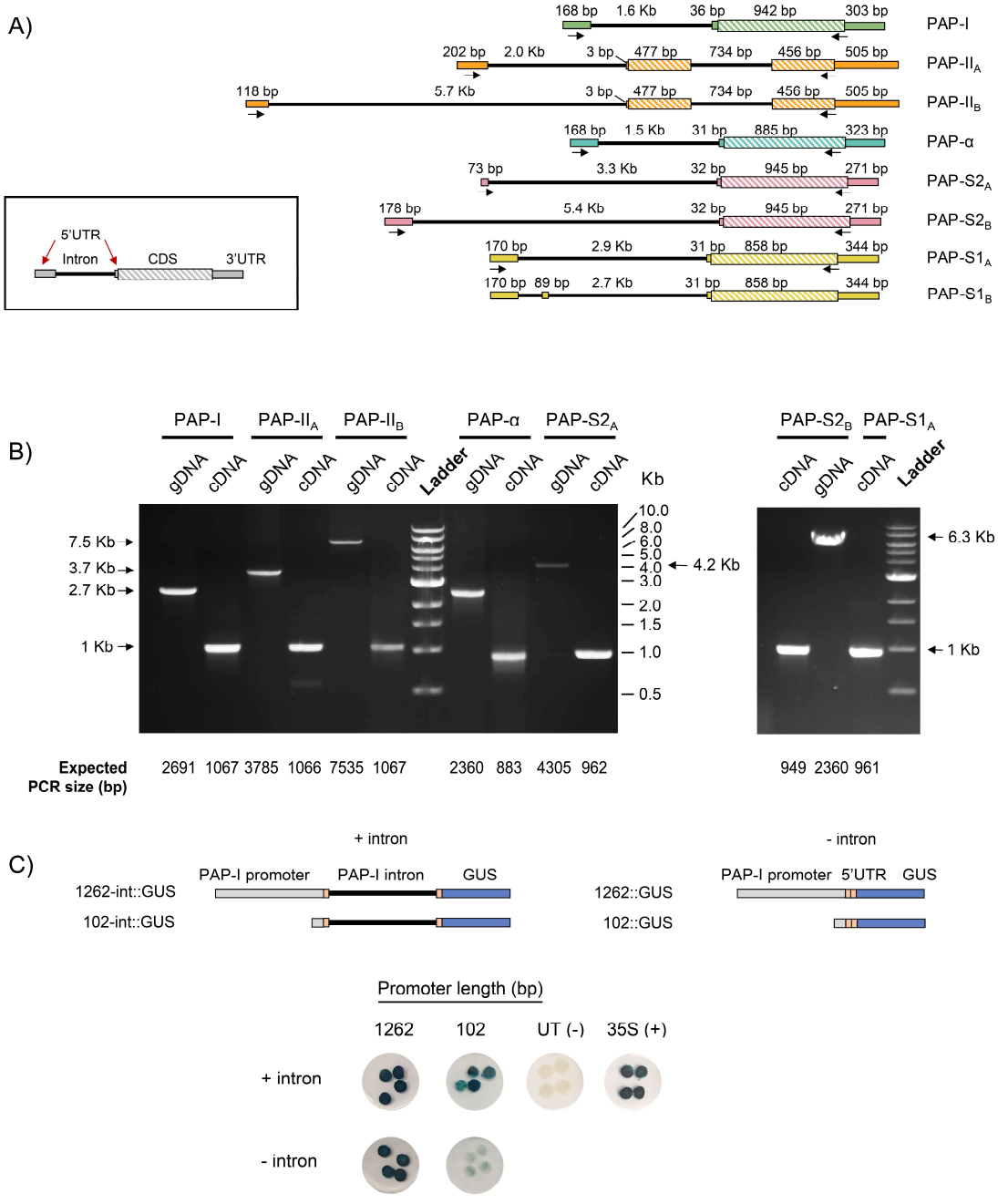



**Figure 18. Legend on next page.**

**Figure 18. Identification of a novel intron in the 5' UTR of PAP genes**. **A)** Gene models of PAP isoforms obtained from annotation with the MAKER pipeline. Arrows indicate primer binding sites used for gene model validation. **B)** Validation of PAP gene models through PCR and RT-PCR from pokeweed genomic DNA (gDNA) and cDNA, respectively. All products were validated by sequencing. **C)** Effect of the 5' UTR intron on PAP-I gene expression. GUS reporter constructs were created with the 1262 bp or 102 bp PAP-I promoter, with or without the 5' UTR intron. Constructs were agroinfiltrated into tobacco and stained for GUS. CaMV 35S = positive control; untransformed (UT) Agrobacterium = negative control. Four independent plants per construct were tested.

long leader introns, all PAP transcripts had potential upstream open reading frames in their 5' UTRs. The majority of transcripts contained a single upstream open reading frame of four codons in length (PAP-I, PAP-α, PAP-$S2_A$, PAP-$S2_B$), while other transcripts contained longer ones (PAP-$II_A$: 21 codons, 17 codons, 11 codons; PAP-$II_B$: 8 codons; PAP-S1: 25 codons).

Where possible, gene models were validated by sequenced RT-PCR and PCR products from pokeweed total RNA and genomic DNA, respectively. As shown in **Figure 18B**, for each gene-specific primer pair, the PCR product size was consistent with that expected from the gene model for both cDNA and genomic DNA. The gene model of the putative novel isoform (PHYAM_012451) is not indicated in **Figure 18** because its 5' and 3' UTRs could not be annotated; this discrepancy may be solved in the future by scaffolding with longer sequence reads. Nonetheless, we report here the validated gene models of all published PAP isoforms and the finding of a novel, conserved feature: a long intron within the 5' UTR.

To investigate if the 5' UTR intron affected PAP gene expression, we created GUS reporter constructs containing the PAP-I proximal promoter (1262 bp) or minimal promoter (102 bp, putative CAAT and TATA boxes only), either with or without the 5' UTR intron. A reporter construct with the 35S CaMV promoter served as the positive control, and untransformed Agrobacterium was the negative control. Agroinfiltration of the constructs into tobacco leaves and subsequent GUS staining revealed that PAP-I promoter constructs with the intron had higher expression than those without (**Figure 18C**). The effect of the intron was most evident for the 102 bp promoter, where presence of the intron increased the level of GUS substantially, relative to the nearly undetectable level of staining without the intron. Based on this preliminary characterization, we hypothesize that the 5' UTR intron, which is a conserved feature of PAP gene models, enhances PAP gene expression.

**Identification of stress-responsive genes in pokeweed**

To gain insight into the response of pokeweed to biotic and abiotic stresses, we identified DEGs from genome-aligned RNA-Seq reads derived from several conditions. Pokeweed plants were treated with JA, SA, PEG, or WND, with ET or WT plants serving as controls for JA/SA and PEG/WND, respectively. EdgeR differential expression results and TPM-normalized expression values of all genes are provided in **Supplementary Data Sheets 4** and **5**. **Figure 19A** shows a heat map of normalized expression of the top DEGs (FDR < 0.001, FC > 4 in at least one pairwise comparison). It total, 3,548 genes were differentially expressed at this threshold. The treatment-specific distribution of DEGs (FDR < 0.05) is shown in **Figure 19B**. The number of DEGs per treatment was as follows: SA (10,310), JA (9,088), PEG (1,549), WND (568); therefore, pokeweed was most responsive to SA and JA, which both mediate pathogen defense, and much less responsive to the abiotic stresses. Interestingly, 58 DEGs (1.6%) were common to all four treatments, including two PAP isoforms (PAP-II and PAP-α). These common DEGs were significantly enriched in the following GO terms (FDR < 0.05): 'proline metabolic process', 'ornithine metabolic process', and 'putrescine biosynthetic process from arginine, using agmatinase'. The term 'rRNA N-glycosylase activity' was also highly enriched (FDR = 0.28), suggesting that PAP contributes to more widespread stress responses in pokeweed than previously known. Furthermore, the finding that not all PAP isoforms were responsive to all treatments suggests that isoform expression is differentially regulated.

To further investigate the unique expression profiles of PAP isoforms and their potential roles in pokeweed, we identified gene clusters from the top DEGs described above. Using DEclust, which extracts statistically significant gene clusters from multi-conditional transcriptome data, 36 clusters were defined (**Supplementary Figure 3**, **Supplementary Data Sheet 6**). Thirteen clusters

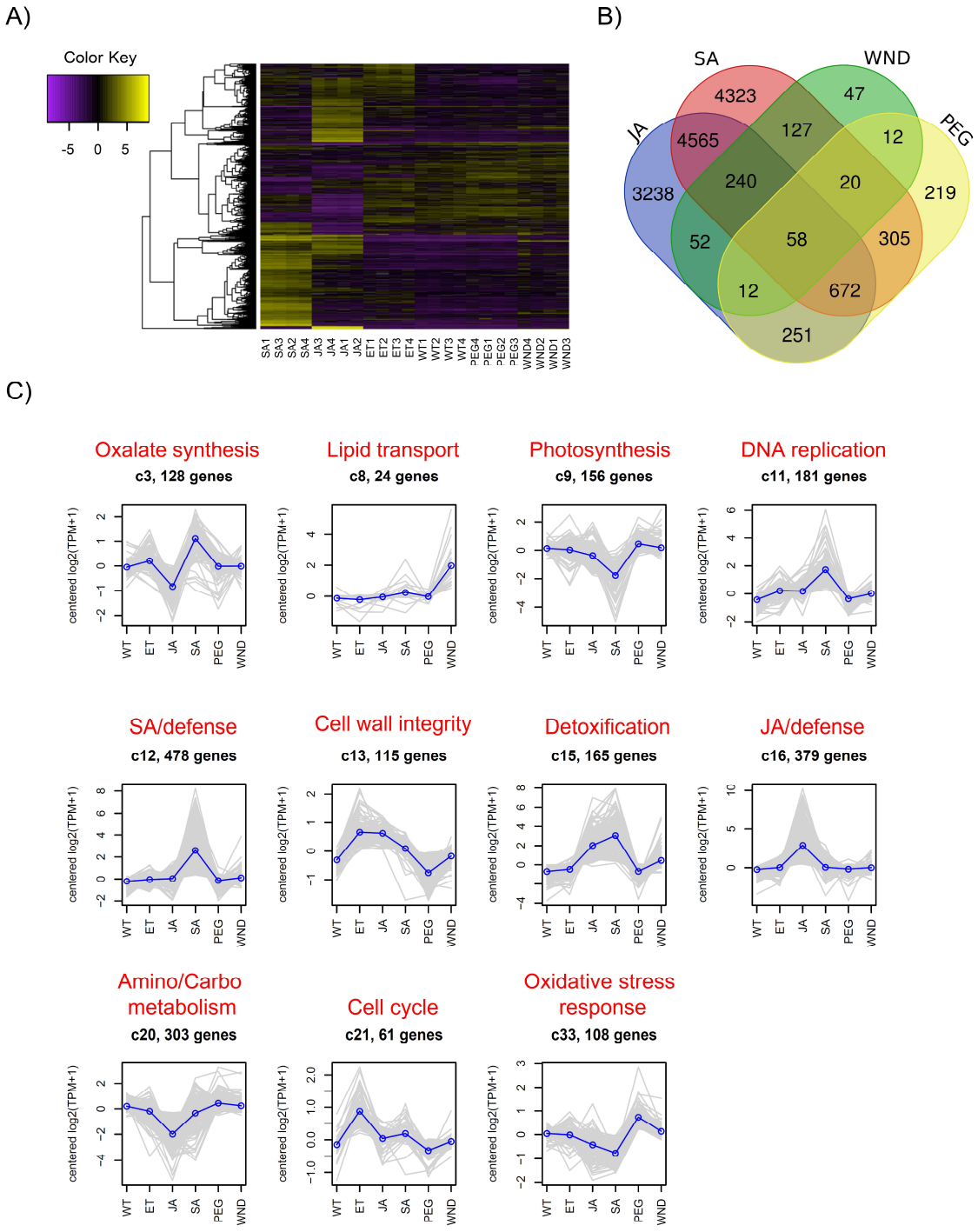



**Figure 19. Legend on next page.**

**Figure 19. Identification of stress-responsive genes in pokeweed. A)** Heat map of normalized expression values ($\log_2$(TPM + 1), median-centered) of the top differentially expressed genes (DEGs; FDR < 0.001, FC > 4 in at least one pairwise comparison). SA, JA, ET, WND, WT, and PEG denote salicylic acid, jasmonic acid, ethanol, wounding, water, and polyethylene glycol treatments, respectively. For each condition, four RNA-Seq libraries were prepared from three independent pokeweed plants (i.e. n=4 pooled biological replicates). **B)** Treatment-specific distribution of DEGs (FDR < 0.05). **C)** Gene clusters having significant functional enrichment. Top DEGs (from A) were clustered with DEclust. For each cluster, the mean expression profile was plotted as a blue line and the biological relevance of enriched GO terms (FDR < 0.05) was summarized in red font.

revealed significant GO term enrichment (FDR < 0.05). **Figure 19C** shows the expression profiles of these functionally enriched clusters, and **Table 15** provides the associated GO terms. Overall, each cluster had a discrete and unified biological theme, indicating successful resolution of co-expressed genes. We summarized the enriched GO terms per cluster into the following biological themes: oxalate synthesis, lipid transport, photosynthesis, DNA replication, SA-mediated signaling/defense, cell wall integrity, detoxification, JA-mediated signaling/defense, amino acid/carbohydrate metabolism, cell cycle regulation, and oxidative stress response. Two PAP isoforms (PAP-S1 and PAP-α) were present in Cluster 16, which was enriched in GO terms comprising JA-associated responses including 'regulation of jasmonic acid mediated signaling pathway', 'response to wounding', and 'terpenoid biosynthetic process'. While Cluster 16 contained JA-upregulated genes, Cluster 20, which included the PAP-S2 isoform, consisted of JA down-regulated genes enriched in the terms 'fructose-bisphosphate aldolase activity' and 'amino acid export'. Overall, these results provide an indication of how pokeweed responds to diverse stresses and situate PAP within key defense pathways.

**Differential regulation of PAP isoform expression**

The individual expression profiles of PAP genes, including pseudogenes, are provided in **Figure 20A**. PAP gene expression changes were also validated through qRT-PCR for all four stress treatments ($R^2$ = 0.8807; **Supplementary Figure 4**). In addition to showing differences in stress-induced expression change, the isoforms varied greatly in abundance. Among protein-coding isoforms, average abundance in TPM across all samples ranged from 146 (PAP-α) to 9911 (PAP-I). PAP-I was the 24th most expressed gene in the plant under normal conditions (WT) and

**Table 15. Significantly enriched GO terms (FDR < 0.05) in pokeweed gene clusters.**

| Cluster | Enriched Terms |
|---|---|
| **3** | • oxaloacetase activity<br>• citramalate lyase activity |
| **4** | • transferase activity, transferring acyl groups other than amino-acyl groups |
| **8** | • lipid transport<br>• lipid binding<br>• anchored component of membrane |
| **9** | • chlorophyll binding<br>• photosystem I<br>• protein-chromophore linkage<br>• pigment binding<br>• photosynthesis, light harvesting in photosystem I<br>• plastoglobule<br>• response to light stimulus<br>• chlorophyll biosynthetic process |
| **11** | • MCM complex<br>• nucleosome<br>• DNA replication initiation<br>• protein heterodimerization activity<br>• nucleosome assembly<br>• cell cycle<br>• DNA binding<br>• THO complex<br>• single-stranded DNA 3'-5' exodeoxyribonuclease activity<br>• embryonic root morphogenesis<br>• cell division<br>• DNA replication |
| **12** | • protein serine/threonine kinase activity<br>• defense response to bacterium<br>• respiratory chain<br>• plasma membrane<br>• NADH dehydrogenase (ubiquinone) activity<br>• ATP synthesis coupled electron transport<br>• ADP binding<br>• response to salicylic acid<br>• signal transduction<br>• ATP binding |
| **13** | • plant-type secondary cell wall biogenesis<br>• lignin catabolic process<br>• hydroquinone:oxygen oxidoreductase activity<br>• oxidoreductase activity, oxidizing metal ions<br>• proteinaceous extracellular matrix<br>• cell wall<br>• apoplast<br>• xylan O-acetyltransferase activity<br>• cellulose synthase (UDP-forming) activity<br>• abscission<br>• copper ion binding<br>• cellulose biosynthetic process<br>• rhamnogalacturonan I side chain metabolic process<br>• extracellular region<br>• xylan biosynthetic process<br>• hydrolase activity, hydrolyzing O-glycosyl compounds<br>• cell wall organization<br>• cellulose synthase complex |
| **15** | • glutathione transferase activity<br>• serine-type endopeptidase inhibitor activity |
| **16** | • terpenoid biosynthetic process<br>• heme binding<br>• iron ion binding<br>• response to wounding<br>• magnesium ion binding<br>• metabolic process<br>• negative regulation of nucleic acid-templated transcription<br>• regulation of jasmonic acid mediated signaling pathway<br>• oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen |
| **19** | • extracellular region |
| **20** | • fructose-bisphosphate aldolase activity<br>• amino acid export |
| **21** | • spindle microtubule<br>• cell division<br>• chromocenter<br>• mitotic spindle assembly checkpoint |
| **33** | • membrane lipid biosynthetic process<br>• regulation of response to reactive oxygen species<br>• chloroplast membrane response to iron ion starvation |

A)

| Gene ID | Isoform | $log_2$FC JA | $log_2$FC SA | $log_2$FC PEG | $log_2$FC WND | Avg. Expr. (TPM) | Cluster |
|---|---|---|---|---|---|---|---|
| PHYAM_020596 | PAP-I | 4.50 | 2.00 | 0.30 | 0.52 | 9911 | 6 |
| PHYAM_028184 | PAP-II | 3.24 | 0.75 | -1.22 | 0.82 | 1834 | 17 |
| PHYAM_021314 | PAP-S1 | 9.62 | 1.68 | -0.65 | -0.47 | 886 | 16 |
| PHYAM_010467 | PAP-S2 | -0.47 | -0.06 | -0.08 | -0.04 | 495 | 20 |
| PHYAM_022058 | PAP-α | 4.87 | 2.53 | -1.20 | 0.91 | 146 | 16 |
| PHYAM_012451 | Novel | 0.21 | 0.19 | -0.72 | -0.44 | 1092 | - |
| PHYAM_008178 | Pseudogene | - | - | - | - | 0 | - |
| PHYAM_008179 | Pseudogene | - | - | - | - | 0 | - |
| PHYAM_010465 | Pseudogene | 12.38 | - | - | - | 57 | 17 |
| PHYAM_010468 | Pseudogene | 1.27 | -0.40 | 0.09 | -0.31 | 8 | - |

B)

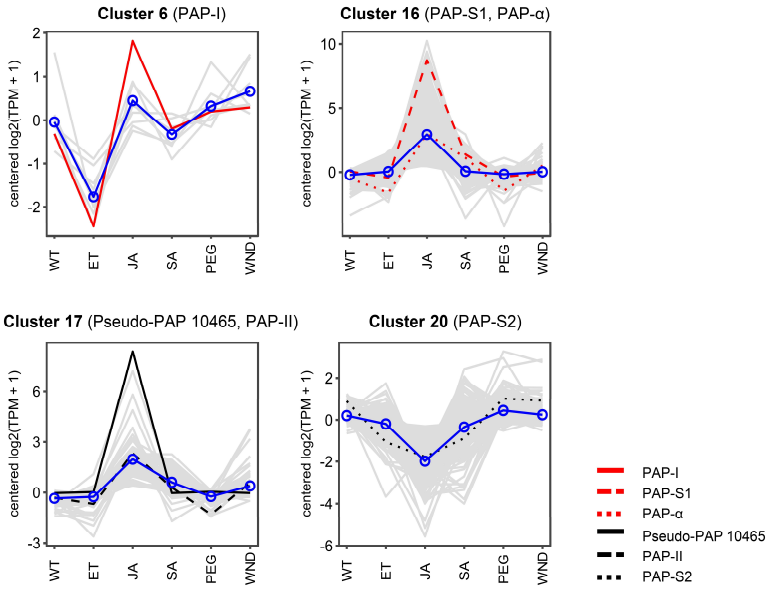



**Figure 20. PAP gene expression profiles. A)** Expression changes of all PAP genes, including pseudogenes, in response to pokeweed stress treatments. Green and red indicate significant up- or down-regulation, respectively. The assigned cluster and average abundance of each isoform are also shown. **B)** Expression profiles of PAP-containing clusters. The profile of the relevant PAP isoform (red or black line) is indicated, as well as the mean expression profile of all genes in the cluster (blue line).

third most expressed upon JA treatment. As expected, only two of the four PAP pseudogenes showed quantitative evidence of transcriptional expression, and their abundances (TPM = 57 and 8) were much less than those of protein-coding isoforms (average TPM = 2,394). Both transcribed pseudogenes were JA-responsive, one of which (PHYAM_010465) was assigned to Cluster 17 with PAP-II (**Figure 20B**). Given that pseudogenes can regulate post-transcriptional expression of functional parental genes, our finding that a PAP pseudogene is co-expressed with a PAP isoform may indicate a novel mechanism by which PAP gene expression is controlled.

PAP-I and PAP-II, the two most abundant isoforms, were assigned to clusters that lacked significant functional enrichment (**Figure 20B**). PAP-I was responsive to both JA and SA. It was assigned to Cluster 6, which included only eight other genes, six of which had annotated homologs: two transcription factors from the HD-ZIP homeobox family (HAT5 and ATHB-7), pathogenesis-related protein STH-21, and the enzymes galactinol synthase 2, strigolactone esterase DAD2, and 1,4-dihydroxy-2-naphthoyl-CoA thioesterase. PAP-II, like PAP-α, was responsive to all stresses; however, the two isoforms were assigned to different clusters (17 and 16, respectively). The mean expression profiles of both clusters were similar, exhibiting a prominent peak with JA treatment. Accordingly, Cluster 17 was enriched in the terms 'response to wounding' (FDR = 0.10) and 'jasmonic acid biosynthetic process' (FDR = 0.23). Assignment of PAP-II and PAP-α to different clusters likely reflects differences in the intensity of their responses to JA ($log_2FC$ = 3.24 and 4.87, respectively) and SA ($log_2FC$ = 0.75 and 2.53, respectively). The putative novel isoform (PHYAM_012451) was most unlike the others. It responded only to PEG, showing a small but significant decrease in expression ($log_2FC$ = -0.72). Because this isoform was not among the top DEGs, it was not included in the clustering analysis; future work will investigate treatments that

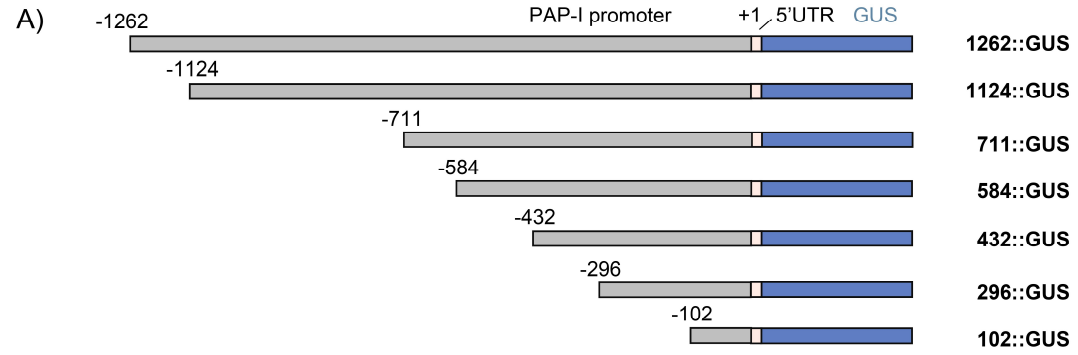
induce its expression. Taken together, these results demonstrate that PAP isoforms have distinct expression profiles, suggesting that they contribute to unique functions in pokeweed.

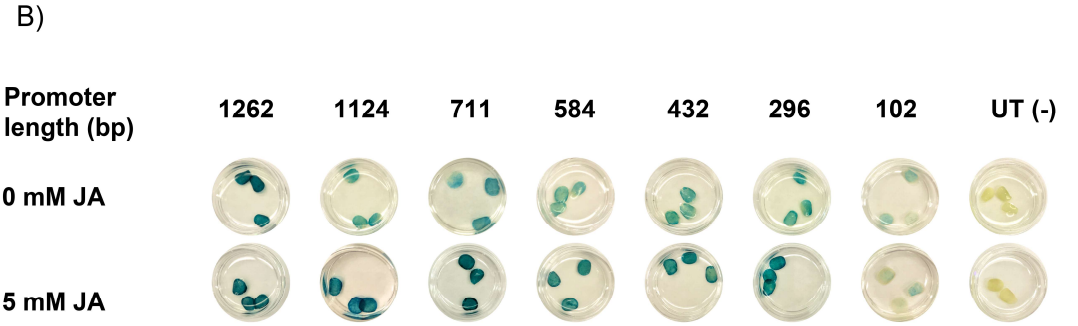
**JA-responsiveness of the PAP-I promoter**

Since PAP-I was highly up-regulated with JA ($log_2FC = 4.5$), we hypothesized that the PAP-I promoter contained CREs to mediate this response. PAP-I promoter fragments (ranging from 1262 bp to 102 bp) were placed upstream of the GUS reporter gene and transiently expressed in tobacco leaves through agroinfiltration (**Figure 21A**). Apart from those expressing 102::GUS, all plants bearing the PAP-I promoter::GUS constructs showed higher GUS activity following JA treatment, relative to controls without JA, as determined through GUS histochemical and fluorometric assays (**Figure 21B** and **21C**). Therefore, our results suggest that the region upstream of the TSS (-296 to -103) is sufficient for the response of PAP-I to JA. As shown in **Figure 21D**, bioinformatic annotation of CREs in this region revealed an element (T/GBOXPINAT2) that binds the master JA signalling regulator MYC (Boter et al., 2004). Additionally, binding sites for TFs of the bHLH, bZIP, and MYB families were present. Specifically, MYB family members bind MYB motifs, while members of bHLH and bZIP families bind the T/GBOXATPIN2 element. Given that pokeweed homologs of several of these TFs were present in JA-associated Cluster 16 (bZIP2, bZIP11, bHLH14, bHLH25, bHLH35, MYB15, MYB44, MYB62, MYB305), the annotated CREs in this region likely contribute to JA-responsiveness of the PAP-I promoter.

**Stress-associated CREs in the promoters of PAP isoforms**

To gain further insight into the function of PAP isoforms, we extended our bioinformatic annotation of CREs to include the proximal promoters of all protein-coding PAP genes, including the putative alternate promoters of PAP-II and PAP-S2 (**Table 16**). All promoters contained TATA



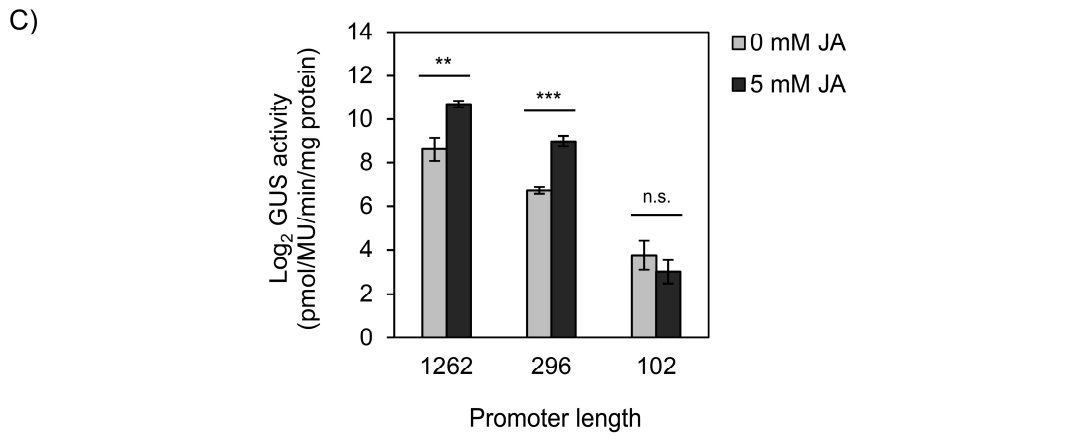





**Figure 21. Legend on next page.**

**Figure 21. Response of the PAP-I promoter to JA. A)** Schematic illustration of PAP-I promoter::GUS constructs. The PAP-I promoter was serially truncated from the 5' end. +1 denotes the transcription start site (TSS). **B)** GUS histochemical assay in tobacco leaves agroinfiltrated with PAP-I promoter::GUS constructs and treated with either 0 mM (mock) or 5 mM JA. Untransformed (UT) Agrobacterium = negative control. Three independent plants per construct were tested. **C)** GUS fluorometric assay in tobacco leaves agroinfiltrated with PAP-I promoter::GUS constructs and treated with either 0 mM (mock) or 5 mM JA. At least four independent plants per constructs were tested. Error bars represent the standard error of the mean (SEM). Comparisons between mock-treated and JA-treated samples were conducted for each promoter construct using two-tailed t-tests, $p < 0.01$. '**' = $p < 0.01$; '***' = $p < 0.001$; 'n.s.' = not significant. **D)** Stress-related CREs in a region (-296 to +1) of the PAP-I promoter. Nucleotide position is indicated on the left, relative to the validated TSS (+1). Sequences of stress-related CREs, along with the CAAT and TATA boxes, are bolded and boxed. Green font indicates a known JA-associated element.

**Table 16. Putative stress-responsive cis-regulatory elements (CREs) in the PAP promoters.**

| CRE | Motif | Response | Copy number in PAP promoter | | | | | | | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | I | $II_A$ | $II_B$ | α | S1 | $S2_A$ | $S2_B$ | |
| ABRELATERD1 | ACGTG | ABA; Dehydration | 2 | 1 | 3 | 1 | - | 1 | 2 | (Simpson et al., 2003) |
| ABREMOTIFAOSOSEM | TACGTGTC | ABA | - | 1 | - | - | - | - | - | (Hobo et al., 1999) |
| ABREOSRAB21 | ACGTSSSC | ABA | 1 | - | - | - | - | - | - | (Marcotte et al., 1989) |
| ABRERATCAL | MACGYGB | ABA; $Ca^{2+}$ | 2 | - | 1 | - | - | - | - | (Kaplan et al., 2006) |
| ACGTATERD1 | ACGT | Dehydration | 6 | 2 | 4 | 4 | - | 2 | 14 | (Simpson et al., 2003) |
| ANAERO1CONSENSUS | AAACAAA | Low oxygen | 1 | 2 | 1 | 1 | - | 1 | 2 | (Mohanty et al., 2005) |
| ANAERO2CONSENSUS | AGCAGC | Low oxygen | - | - | 2 | - | - | - | - | (Mohanty et al., 2005) |
| BIHD1OS | TGTCA | Disease | 1 | 2 | 4 | 6 | 5 | 4 | 6 | (Luo et al., 2005) |
| CACGTGMOTIF | CACGTG | Defense-related | - | - | 2 | - | - | - | - | (Gu et al., 2002) |
| CAREOSREP1 | CAACTC | GA | 2 | 1 | 1 | 1 | 1 | 1 | - | (Sutoh and Yamauchi, 2003) |
| CCAATBOX1 | CCAAT | Heat stress | 4 | 6 | 2 | - | 1 | - | 2 | (Rieping and Schöffl, 1992) |
| DPBFCOREDCDC3 | ACACNNG | ABA | - | 2 | 1 | - | 2 | 0 | 1 | (Kim et al., 1997) |
| DRE2COREZMRAB17 | ACCGAC | ABA; Dehydration | - | 2 | - | - | 1 | 1 | - | (Kizis and Pagès, 2002) |
| DRECRTCOREAT | RCCGAC | Cold; Dehydration; High salt | - | 3 | - | - | 1 | 1 | - | (Dubouzet et al., 2003) |
| EBOXBNNAPA | CANNTG | Cold | 6 | 12 | 8 | 8 | 2 | 6 | 12 | (Agarwal et al., 2006) |
| ELRECOREPCRP1 | TTGACC | Pathogen; SA; Wounding | - | - | 2 | - | - | - | - | (Laloi et al., 2004) |
| GADOWNAT | ACGTGTC | GA | - | 1 | - | - | - | - | - | (Ogawa et al., 2003) |
| GARE1OSREP1 | TAACAGA | GA | - | - | 1 | 1 | - | - | - | (Sutoh and Yamauchi, 2003) |
| GARE2OSREP1 | TAACGTA | GA | - | - | - | - | - | - | 2 | (Sutoh and Yamauchi, 2003) |
| GAREAT | TAACAAR | GA | 2 | 1 | 1 | 2 | 2 | 2 | - | (Ogawa et al., 2003) |
| GT1CONSENSUS | GRWAAW | Light-inducible; SA | 17 | 16 | 16 | 12 | 13 | 14 | 14 | (Buchel et al., 1999) |
| GT1GMSCAM4 | GAAAAA | Pathogen; High salt | 6 | 6 | 5 | 3 | 4 | 5 | 2 | (Park et al., 2004) |
| LTRE1HVBLT49 | CCGAAA | Cold | 1 | 1 | - | - | - | - | 1 | (Alison Dunn et al., 1998) |
| LTRECOREATCOR15 | CCGAC | Cold; Dehydration | 1 | 6 | 1 | - | 2 | 2 | - | (Kim et al., 2002) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| MYB1AT | CANNTG | ABA; Dehydration | 5 | 4 | 2 | 6 | 6 | 6 | 3 | (Abe et al., 2003) |
| MYB1LEPR | GTTAGTT | Defense-related | - | 1 | - | - | 1 | 1 | 1 | (Chakravarthy et al., 2003) |
| MYB2CONSENSUSAT | YAACKG | ABA; Dehydration | - | 1 | 1 | - | - | 1 | - | (Abe et al., 2003) |
| MYBCORE | CNGTTR | ABA; Dehydration | - | 3 | 3 | 1 | 1 | 1 | 1 | (Abe et al., 2003) |
| MYBGAHV | TAACAAA | GA | 1 | 1 | 1 | 1 | 1 | 1 | - | (Gubler et al., 1995) |
| MYCATERD1 | CATGTG | Dehydration | - | - | 1 | - | - | 1 | 1 | (Simpson et al., 2003) |
| MYCATRD22 | CACATG | Dehydration | - | - | 1 | - | - | 1 | 1 | (Simpson et al., 2003) |
| PREATPRODH | ACTCAT | Hypoosmolarity | - | 2 | 1 | 2 | - | - | 1 | (Satoh et al., 2004) |
| PYRIMIDINEBOXHVEPB1 | TTTTTTCC | ABA; GA | 1 | - | - | - | 1 | - | 1 | (Cercós et al., 1999) |
| PYRIMIDINEBOXOSRAMY1A | CCTTTT | Sugar repression; GA | 4 | - | - | 2 | 1 | 1 | - | (Mena et al., 2002) |
| RYREPEATBNNAPA | CATGCA | ABA | - | 1 | - | 2 | 5 | 4 | - | (Ezcurra et al., 1999) |
| SURECOREATSULTR11 | GAGAC | Sulphur | 4 | 1 | 2 | 2 | 1 | 1 | 1 | (Maruyama-Nakashita et al., 2005) |
| T/GBOXATPIN2 | AACGTG | JA | 2 | - | - | 1 | - | - | 1 | (Boter et al., 2004) |
| WBOXATNPR1 | TTGAC | SA | 2 | 2 | 5 | 6 | 4 | 3 | 3 | (Yu et al., 2001) |
| WBOXNTCHN48 | CTGACY | Defense-related | - | - | 2 | - | - | - | - | (Yamamoto et al., 2004) |
| WBOXNTERF3 | TGACY | Wounding | 4 | 4 | 8 | 5 | 4 | 3 | 3 | (Nishiuchi et al., 2004) |

boxes -35 to -25 bp upstream of the putative TSS, consistent with other eukaryotic promoters. We also identified CREs associated with diverse biotic and abiotic stresses, such as T/GBOXATPIN2 (JA), W-boxes (SA), ABRE motifs (ABA), CCAAT boxes (heat stress), GARE motifs (gibberellic acid, GA), and MYB motifs (drought). Although some CREs were present in all PAP promoters (e.g. EBOXBNNAPA, GT1CONSENSUS, MYB1AT, and WBOXATNPR1), most elements differed in abundance and distribution. For instance, most ABRE motifs were absent in the PAP-α, PAP-S1, and PAP-$S2_A$ promoters, while T/GBOXATPIN2 was only present in PAP-I, PAP-α, and PAP-$S2_B$. Other elements were unique to a single promoter, such as ELRECOREPCRP1 (PAP-$II_B$), GADOWNAT (PAP-$II_A$), and WBOXNTCHN48 (PAP-$II_B$). The two PAP-II promoters had only 50.5% sequence identity to each other, and the two PAP-S2 promoters had only 47.2% identity. Therefore, stimuli-dependent promoter selection in these two isoforms may lead to the transcription of distinct populations of mRNAs differing only in their 5' UTRs. Overall, differences in the abundance and distribution of CREs in PAP promoters suggest that the isoforms have distinct roles and that control of expression can be fine-tuned by isoforms with more than one promoter.

**4.5 Discussion**

Here, we have presented the first de novo assembled draft genome of pokeweed and an annotation of protein-coding genes. We also identified clusters of co-expressed genes by integrating RNA-Seq data from several pokeweed stress treatments. We found that PAP isoforms localized to multiple clusters, with some isoforms clustering together, and functional enrichment analysis suggested distinct biological relevance of isoforms. Validation of PAP gene models led to the discovery of a long intron within the 5' UTR. The sequence of the intron varied for each

isoform, but its presence was consistent. For PAP-I, the intron enhanced gene expression of promoter reporter constructs in tobacco. Finally, we confirmed JA-responsiveness of the PAP-I promoter in tobacco and identified a region that mediates this response. This region, as well as the proximal promoters of all PAP isoforms, contained CREs associated with stress.

**Evaluation of assembly and annotation metrics of the pokeweed genome**

The pokeweed genome was sequenced exclusively as paired-end reads from a single short-insert library (83X coverage), and the resulting de novo assembly accounted for 74% of the expected size. We estimated the genome of pokeweed to be 1.3 Gb, in agreement with the previously reported value of 1.5 Gb obtained by Feulgen microdensitometry (Bennett, 2000). However, the assembly was highly fragmented (~850,000 total scaffolds, with 70,834 scaffolds ≥ 1 Kb; N50 = 42.5 Kb). Available plant genomes have N50 values ranging from $10^3$ to $10^8$ bp (Veeckman et al., 2016). Nevertheless, the contiguity of the pokeweed genome assembly is comparable to other de novo assembled genomes derived from paired-end sequencing of short-insert libraries (Polashock et al., 2014; Van Hoeck et al., 2015). An assembly can also be highly complete but fragmented, as seen for a petunia species whose assembly accounted for 93% of the expected 1.1 Gb genome but had an N50 value of 17.9 Kb (Zhuang and Tripp, 2017). We acknowledge that more advanced sequencing would improve our genome assembly, however the main goal of our draft genome was to identify PAP genes and their proximal promoters.

Annotation metrics indicate that our assembly sufficiently captured the protein-coding gene content of pokeweed. BUSCO completion scores for the genome assembly and gene set were 84% and 76%, respectively. Since the CoreGF score of the gene set was high (98%), and this metric is less stringent in terms of species conservation (Veeckman et al., 2016), we attribute

missing BUSCOs more so to the divergence of pokeweed from model plants than incomplete assembly. Indeed, BUSCO scores are known to reflect both assembly contiguity and evolutionary history of the species under study (Simão et al., 2015). Furthermore, a similar BUSCO genome completion score (77%) was obtained for a non-model plant, a seagrass species, whose assembly metrics were similar to pokeweed: 71% assembled and N50 = 36.7 Kb (Lee et al., 2016). Pokeweed annotation metrics are also consistent with standards set by MAKER developers (Campbell et al., 2014a), who consider a genome to be well-annotated if 90% of its annotations have an AED less than 0.5 and over 50% of its proteome contains a recognizable protein domain. Furthermore, annotation of a plant genome with MAKER is expected to identify at least 20,000-40,000 genes. The pokeweed gene set meets these criteria: ~30,000 genes, 99% of which have an AED less than 0.5, and 73% of the proteome contains a Pfam domain.

**Relevance of pokeweed-specific orthogroups in plant defense**

Pokeweed-specific orthogroups were enriched in the GO terms ‘far-red light signaling pathway’ and ‘negative regulation of defense response’, with the involved genes annotated as isoforms of FAR1-Related Sequence and CPR30, respectively. FAR1 is a TF involved in a variety of processes relating to growth and development, including light signal transduction, circadian clock and flowering time regulation, chlorophyll biosynthesis, starch synthesis, and ABA responses (Ma and Li, 2018). FAR1, together with the light-signalling factor Far-Red Elongated Hypocotyl 3 (FHY3), regulates plant defense by integrating chlorophyll biosynthesis and SA signalling in *Arabidopsis* (Wang et al., 2016). The F-box protein CPR30 also modulates defense in *Arabidopsis* (Gou et al., 2009). Plants deficient in CPR30 showed resistance to pathogen infection and induction of defense-related gene expression. Despite the dependence of both FAR1

and CPR30 on SA in *Arabidopsis*, we did not identify any significant changes in expression of these genes in any of the pokeweed stress treatments. This may reflect the fact that our stress treatments were short-term on wild-type plants and did not simulate a mutant condition. Specifically, both *cpr30* and *fhy3 far1* plants had dwarf phenotypes, indicative of disruption to wider pathways of growth and development. Our identification of pokeweed-specific orthogroups enables future comparison with the agricultural crop plants used in our analysis, to identify defense strategies unique to pokeweed.

**Annotation of PAP isoforms in pokeweed**

Through PAP isoform annotation, we confirmed the existence of the following previously identified isoforms: PAP-I (Irvin, 1975), PAP-II (Irvin et al., 1980), PAP-S1 (Honjo et al., 2002), PAP-S2 (Honjo et al., 2002), and PAP-α (Kataoka et al., 1992). We did not find evidence for PAP-H (Park et al., 2002), PAP-C (Barbieri et al., 1989), PAP-R (Bolognesi et al., 1990), or PAP-III (Rajamohan et al., 1999) in the scaffold assembly. However, a sequence having high identity (95%) and coverage (97%) with the cDNA clone of PAP-H was identified in the edge assembly through BLAST-N. Therefore, a gene for PAP-H likely exists in pokeweed, but it was excluded from the scaffold assembly since it was lower in coverage than an alternative path through that region. PAP-H, with 67% identity to PAP-I at the protein level, was purified from *Agrobacterium rhizogenes*-transformed hairy roots of pokeweed (Park et al., 2002). It is secreted as part of the root exudates and hypothesized to contribute to the inhibition of soil-borne microbe infection. We did not find evidence for PAP-C or PAP-R, which were originally purified from pokeweed cell cultures and roots, respectively. Both PAP-C and PAP-R have N-terminal sequences that are identical to PAP-I; additionally, the three isoforms are highly similar in terms of amino acid

composition, molecular weight (~29 kDa), and pI value (~9.5) (Barbieri et al., 1989; Bolognesi et al., 1990). Based on a lack of genomic support and high biochemical similarity, we suggest that reports of PAP-C, PAP-R, and PAP-I refer to the same isoform. Minor differences in biochemical properties could be explained by experimental variability, different post-translational modifications, or allelic diversity that cannot be resolved through de novo assembly. Finally, we did not identify genomic evidence in support of PAP-III, originally purified from late summer leaves. Through BLAST-P, we determined that the sequence of PAP-II, from early summer leaves, is 95% identical to the sequence of PAP-III at the protein level. Mismatches in the alignment resulted from ambiguous nucleotides (x) in the PAP-III sequence, owing to lysine methylation that enabled protein crystallization in the original report (Kurinov and Uckun, 2003). Therefore, PAP-II and PAP-III have the same amino acid sequence. It is curious that the two proteins have different levels of antiviral activity and separate as distinct peaks in ion-exchange chromatography (Rajamohan et al., 1999). We hypothesize that differences in PAP-II and PAP-III arise from post-translational modifications that affect enzymatic activity. Further support for this idea comes from the finding that a rare form of N-glycosylation exists in PAP seed isoforms, and this modification is thought to contribute to their high cytotoxicity (Hogg et al., 2015; Islam et al., 1991).

**Integration of abiotic and biotic stress responses in pokeweed**

Clustering analysis enabled the identification of genes sharing a similar expression profile. We identified 36 gene clusters, 13 of which showed significant functional enrichment. While clusters associated with SA or JA were expected since these hormones have well-established roles in plant defense, we focus here on clusters revealing potential cross-talk of key biotic and abiotic stress responses in pokeweed.

The ‘cell wall integrity’ cluster included several enzymes involved in the synthesis of cellulose and lignin, which are critical components of the cell wall. Lignin creates a physical barrier against pathogens and makes plant cells more difficult for insect herbivores to penetrate and digest (Liu et al., 2018). Multiple laccases, enzymes required for lignin polymerization, were present in this cluster. A laccase was shown to mediate broad-spectrum pathogen resistance in cotton by integrating the phenylpropanoid pathway, JA biosynthesis, and balance of JA-SA defense responses (Hu et al., 2018). In pokeweed, an increase in laccase activity during Mn treatment is thought to contribute to heavy metal tolerance by reducing the level of toxic reactive oxygen species (Gao et al., 2012). Also present in this cluster were fasciclin-like arabinogalactan proteins (FLAs). These are cell surface adhesion proteins that enable cell expansion during salt stress (Shi, 2003). Consistent with their presence, the cluster expression profile showed down-regulation with PEG, indicating sensitivity to osmotic changes. Finally, the inclusion of Defective in Induced Resistance 1 (DIR1) in this cluster, an essential signalling component of systemic acquired resistance (Champigny et al., 2013; Maldonado et al., 2002), provides further indication of integrated biotic and abiotic stress responses in pokeweed.

In the ‘lipid transport’ cluster, we identified several genes encoding non-specific lipid transfer proteins (LTPs). These are small, basic, cysteine-rich proteins that localize to the apoplast and transport various lipids. LTPs are involved in the synthesis of lipid barrier polymers, such as cuticular wax, and their expression is induced by abiotic stress (Salminen et al., 2016). LTPs also appear to have a role in JA signalling. In barley, JA biosynthesis enzymes were found to produce a covalent adduct consisting of an LTP and reactive oxylipin (Bakan et al., 2006). Furthermore, exogenous application of an LTP-JA complex to grapevine produced a higher antifungal response than either component individually (Girault et al., 2008). This gene cluster had a relatively flat

expression profile apart from a spike at wounding treatment, perhaps reflecting involvement of LTPs in specific activation of the wound-response branch of JA signalling (Wasternack and Song, 2017).

The 'oxalate synthesis' cluster included multiple genes encoding Petal Death Protein (PDP). PDP is an isocitrate lyase in senescent flower petals that catalyzes the conversion of oxaloacetate to acetate and oxalate (Lu et al., 2005). In *Medicago truncatula*, calcium oxalate plays a role in defense against chewing insects, with a clear feeding preference observed for plants defective in calcium oxalate production (Korth, 2006). Authors also found that calcium oxalate crystals were abrasive to insects during feeding and interfered with digestion. Oxalate, as oxalic acid, is used to chelate and detoxify heavy metals in plants (Yang et al., 2005). Notably, pokeweed has an intrinsically high oxalate content that is sufficient for chelating Mn at high concentrations (Dou et al., 2009). In this gene cluster, we also identified a heavy metal-associated isoprenylated plant protein. These metallochaperones have a known role in cadmium detoxification (Tehseen et al., 2010). With pathogen resistance and heavy metal tolerance being the most cited applications of pokeweed, our results provide insight into genes that may mediate cross-talk between these important activities.

**Biological relevance of PAP in pokeweed**

Our identification of genes co-expressed with different PAP isoforms provides a foundation for exploring their regulation in the plant and biological relevance. Two PAP-containing clusters (16 and 20) revealed significant GO term functional enrichment. Cluster 20, which contained PAP-S2 and showed down-regulation with JA, was enriched in terms relating to glycolysis and amino acid transport. The respective genes were annotated as fructose-bisphosphate

aldolases (FBAs) and WAT1-related proteins. WAT1 is a vacuolar transporter that promotes indole metabolism and transport in *Arabidopsis* (Denancé et al., 2013; Ranocha et al., 2010). Indole is important in plant defense as a herbivore-induced volatile priming signal (Erb et al., 2015) and precursor for secondary metabolites (Lee et al., 2015). This cluster was also enriched in FBA genes, which function in glycolysis and gluconeogenesis. Co-enrichment of genes involved in carbohydrate and indole metabolism in this cluster suggests recruitment of the shikimate pathway, which provides a route to aromatic amino acid biosynthesis from chorismate (Parthasarathy et al., 2018). Specifically, the glycolysis intermediate phosphoenol pyruvate is used as input for the shikimate pathway, and indole is synthesized from tryptophan. FBA is also involved in the Calvin cycle of photosynthesis. Interestingly, this cluster included genes encoding pentatricopeptide repeat-containing (PPR) proteins that participate in RNA editing in the chloroplast (PCMP-H61, PCMP-H51/CRR28), an activity required for accumulation of the NADH dehydrogenase-like complex of the photosynthetic electron transport chain (Okuda et al., 2007, 2009). PPRs are a large family of sequence-specific RNA-binding proteins involved in multiple aspects of RNA metabolism. Taken together, these results demonstrate PAP-S2 co-expression with genes involved in key processes of carbohydrate and amino acid metabolism, including other proteins that interact directly with RNA.

Cluster 16, which included PAP-S1 and PAP-α, contained JA-upregulated genes and had corresponding GO enrichment in JA signalling and terpenoid biosynthesis. Like Cluster 20, this cluster also indicated PAP co-expression with genes involved in metabolism, specifically with the presence of BZIP11, a sucrose-regulated transcription factor that controls amino acid and carbohydrate metabolism and is integrated with a wider plant growth regulatory network (Hanson et al., 2008; Ma et al., 2011). Another gene in this cluster, annotated as Enhanced Downy Mildew

2 (EDM2), is an interesting candidate that links regulation of genes with long introns and pathogen defense, both of which are relevant to PAP. EDM2 is a chromatin regulator that promotes normal 3' distal polyadenylation of transcripts from genes containing intronic heterochromatin, the latter arising from methylated transposons or repeats associated with long introns (Duan et al., 2017; Lei et al., 2013). EDM2 is also required for pathogen resistance in *Arabidopsis* by regulating transcript accumulation of an NB-LRR disease resistance (R) gene (Eulgem et al., 2007).

Clusters 6 and 17, which contained PAP-I and PAP-II, respectively, did not have significant functional enrichment (FDR < 0.05). However, assignment of these isoforms to independent clusters reinforces their distinct regulation in the plant. The PAP-II cluster contained 92 genes and was enriched in the GO terms 'wounding' (FDR = 0.10) and 'JA biosynthetic process' (FDR = 0.23). The PAP-I cluster contained only nine genes, including two transcription factors from the HD-ZIP homeobox family that regulate plant growth and leaf development in response to abiotic stresses (Aoyama et al., 2007; Söderman et al., 1996). Other genes in the PAP-I cluster support an integration with plant growth responses. This includes DAD2, an esterase that mediates strigolactone signaling (Hamiaux et al., 2012). Strigolactone is a hormone involved in branching and symbiotic interactions with soil microbes (Marzec, 2016). Also present is a gene annotated as DHNAT1, encoding an enzyme (1,4-dihydroxy-2-naphthoyl-CoA thioesterase 1) involved in the synthesis of phylloquinone (Widhalm et al., 2012). Phylloquinone is required for Photosystem 1 stability (Wang et al., 2017) and is synthesized from chorismate, the final product of the shikimate pathway described above. These findings support the idea that PAP expression regulation is tied to the broader metabolic state of the plant. Specifically, our results indicate co-expression of PAP isoforms with genes involved in amino acid and carbohydrate metabolism, suggesting a link to wider nutrient-sensing networks. Therefore, the ribosome-inactivating activity

of PAP may contribute to and/or be affected by large-scale reprogramming in response to plant stress. This hypothesis is strengthened given the direct antiviral activity of PAP, which could potentially mediate a trade-off between plant growth and defense in pokeweed.

**Mechanisms of gene expression regulation by leader introns**

Investigation of PAP gene models led to the discovery of a long intron in the 5' UTR of all PAP isoforms. The sequence of the intron was different for each isoform, but its presence was conserved. We also provided evidence to support a functional role for the intron, as the PAP-I intron enhanced reporter gene expression in tobacco. Introns can influence gene expression in both plants and animals, particularly leader introns (Laxa, 2017; Shaul, 2017). The precise mechanism of intron-mediated enhancement is not known, but hypotheses at both the transcriptional and translational levels have been proposed. Leader introns may enhance transcription by creating a favourable zone for transcription initiation; that is, a region: i) devoid of nucleosomes, ii) marked by activating histone modifications, or iii) associated with a factor that recruits transcriptional machinery (Gallegos and Rose, 2015). Additionally, splicing signals in the leader intron may affect mRNA processing, export, or decay, thereby affecting translation (Gallegos and Rose, 2015; Laxa, 2017). In contrast to intron-mediated enhancement, which requires the intron to be in its native orientation and position, a leader intron may act in other ways to enhance expression. For example, the intron may function as a classical enhancer by containing CREs (Kim et al., 2006) or as an alternate promoter (Morello et al., 2002, 2006; Qi et al., 2007).

Our preliminary characterization allowed us to conclude that the PAP-I leader intron enhanced expression. Interestingly, we observed that enhancement was greater when the intron was paired with the minimal PAP-I promoter than with the proximal promoter. There may be a

limit to intron-associated enhancement of gene expression, particularly for promoters that drive high expression on their own. For example, when paired with its native weak promoter, the leader intron of *Arabidopsis AtMHX* increased expression by 270-fold, compared to 3-fold with the strong CaMV 35S promoter (Akua et al., 2010). Alternatively, the enhancement provided by an intron may simply be more detectable in the presence of a weak promoter. Future work will investigate the mechanism by which the leader intron enhances PAP expression.

**Transcriptional control of PAP isoform expression**

The promoters of all PAP isoforms contained CREs associated with diverse abiotic and biotic stresses, suggesting that PAP is broadly implicated in plant defense. Since PAP genes were most responsive to JA in our study, we aimed to identify CREs that could mediate this response. Promoter truncation constructs of PAP-I revealed that a region close to the TSS (-296 to -103) was sufficient for JA-responsiveness. This result agrees with a previous finding that CREs closer to the TSS tend to have a greater influence on transcription (Zou et al., 2011). The -296 to -103 region of PAP-I contains binding motifs for transcription factors from bHLH, bZIP, and MYB families, which were found to be overrepresented in JA-responsive promoters (Hickman et al., 2017). Additionally, this region contains a T/GBOXATPIN2 element, which binds the master JA signalling regulator MYC (Boter et al., 2004). Mutation of this element abolished JA-responsiveness of genes in tomato, *Arabidopsis*, and barley (Boter et al., 2004; Rouster et al., 1997).

Analysis of JA-associated CREs in the promoters of other PAP isoforms revealed that the T/GBOXATPIN2 element was also present in PAP-α and PAP-$S2_B$, but not in PAP-S1, which was most JA-responsive. The lack of this element in some promoters of JA-responsive isoforms may

be compensated by the presence of W-boxes; this element binds WRKY TFs, which are primarily SA-responsive (Dong et al., 2003). However, a substantial fraction of WRKYs (at least 30%) are JA-responsive in *Arabidopsis* (Schluttenhofer et al., 2014), and we identified several that were up-regulated with JA in pokeweed (homologs of WRKY3, WRKY4, WRKY22, WRKY23, WRKY24, WRKY33, WRKY40, WRKY41, WRKY49, WRKY70, and WRKY75). Combined with the knowledge that synergistic activation of gene expression can occur in the presence of both SA and JA (Mur, 2005), we hypothesize that W-boxes in the promoters of PAP isoforms contribute to their JA-responsiveness. Although the promoter of PAP-S2 contained several W-boxes, PAP-S2 expression decreased slightly but significantly with JA and was unresponsive to other treatments. It is possible that PAP-S2 has a different temporal expression profile than the other isoforms and is more responsive outside of the 24h time-point we investigated. Consistent with this hypothesis, an RNA-Seq time-course of the JA response in *Arabidopsis* revealed diverse expression patterns over the first 16h following treatment, including distinct early and late responses (Hickman et al., 2017). In addition to CREs associated with JA, we identified CREs associated with the hormones SA, ABA, and GA. JA and SA have well-established roles in plant defense against pathogens and insect herbivores, while ABA contributes to the resistance of abiotic stresses such as drought, salinity, cold, and heat stress (Verma et al., 2016). GA, through cross-talk with ABA pathways, helps mediate the balance between dormancy and plant maturation during stress (Verma et al., 2016). Importantly, we identified differences in the CREs of PAP promoters, suggesting that the isoforms mediate individual responses to hormones.

The draft genome of pokeweed has provided new information on how PAP expression is controlled. The presence of diverse stress-responsive CREs in the promoters of PAP genes, combined with their distinct expression profiles, provides a foundation to examine the role of PAP

isoforms in pokeweed. One-fifth of land plants synthesize RIPs, and these genes are often up-regulated during stress. Our study contributes to the growing body of evidence illuminating RIPs as important components of plant response to environmental change.

### 4.6 Acknowledgement

The authors are grateful to Dr. Paul Kron, Department of Integrative Biology, University of Guelph, for genome size analysis of pokeweed.

# CHAPTER 5:

# Discussion and future directions

### 5.1 Generation of novel bioinformatic resources for pokeweed

By reporting the pokeweed miRNA transcriptome, mRNA transcriptome, and assembled genome, we have produced the first large-scale sequence resources for a representative member of the Phytolaccaceae plant family. Other genomes of plants that express RIPs have been sequenced, but only a few studies have described RIPs at the genomic level. These include castor bean (Chan et al., 2010), rice (Jiang et al., 2008; Wytynck et al., 2017), bitter gourd (Urasaki et al., 2017), and physic nut (Wu et al., 2015). To the best of our knowledge, we are the only group to apply RNA-seq methods to investigate RIP expression and to identify a small RNA that post-transcriptionally targets a RIP mRNA. Study of pokeweed has primarily been limited to its synthesis of PAP and a handful of other products with biomedical relevance. However, given the broad tolerance of this plant to environmental stress, in its resistance to both pathogens and heavy metal toxicity, we anticipate that our bioinformatic resources will aid discovery of other important genes. PAP has never been knocked out in pokeweed, so it remains unknown whether its endogenous expression is essential for pathogen resistance. It is also unknown whether PAP interacts with other defense proteins in pokeweed. Therefore, the relevance of pokeweed extends beyond PAP.

To this end, sequence comparisons have revealed that pokeweed is distinct relative to other plants for which information is available. In Neller et al. (2016), de novo assembly of the pokeweed mRNA transcriptome from leaf tissue treated with or without JA identified 89,682 pokeweed transcripts (expression $\geq$ 1 FPKM). Only 38% of these were annotated as proteins through BLASTx homology searches, and of 434 JA-responsive transcripts detected in only JA-treated samples, 16% could be annotated by BLASTx. We noted that a fraction of these non-protein coding mRNA transcripts could represent miRNA precursors. However, in Neller et al. (2018), using paired small RNA samples isolated from the same plants as in Neller et al. (2016), we

predicted only 582 miRNAs from pokeweed leaf tissue. Of these, 24 had been previously identified (i.e. conserved) in other plants, and 2 of 145 JA-responsive miRNAs were conserved. In Neller et al. (2019), we annotated 29,773 de novo assembled genes using the pokeweed transcriptome and proteomes from available Caryophyllales. Although 73% of genes contained a conserved protein domain, only 64% had significant hits in the SwissProt database. Furthermore, orthogroup analysis revealed that eighteen orthogroups (116 genes) were pokeweed-specific relative to the other Caryophyllales used for annotation. Taken together, analysis at both transcriptomic and genomic levels indicates that pokeweed encodes novel sequences.

### 5.2 Insight into stress responses in pokeweed

To gain insight into pokeweed stress tolerance, we focussed mainly on the effect of JA. This plant hormone is associated with defense against insects, viruses, and necrotrophic pathogens (Carr et al., 2019; Wasternack and Feussner, 2017). It also mediates tolerance to abiotic stresses, including heavy metal toxicity (Foroughi et al., 2014; Singh and Shah, 2014; Yan et al., 2013). Therefore, JA is likely to play a role in coordinating agriculturally relevant traits in pokeweed. In Neller et al. (2016), JA-induced transcripts were highly enriched in genes involved in stress and defense responses. Among the most upregulated transcripts were those annotated as intracellular ribonuclease LX, bark storage protein A, and defensin-like protein. The former two genes have previously been associated with nutrient recycling in senescent plants (Coleman et al., 1991; Lers et al., 2006). We also determined that JA-responsive pairs of natural antisense transcripts (NATs) were enriched in nitrogen assimilation. These results suggest that pokeweed responds to JA in part by allocating resources toward nutrient mobilization. In Neller et al. (2018), we investigated miRNA control of the JA response in pokeweed. JA-responsive miRNA/target pairs were

associated with JA biosynthesis and signalling, as well as pathways involving SA, ABA, and classic growth hormones (auxin, cytokinin). These genes may represent key points of integration to balance abiotic and biotic stress responses, as well as wider modulation of plant growth versus defense.

In Neller et al. (2019), we extended our investigation to include additional treatments, namely SA, PEG, and wounding. Based on the number of DEGs per treatment, pokeweed was most responsive to SA and JA, both of which mediate pathogen defense. Fewer DEGs were observed in response to abiotic stress treatments. Through co-expression analysis of the top DEGs, we identified significant clusters enriched in the following biological themes: oxalate synthesis, lipid transport, photosynthesis, DNA replication, SA-mediated signaling/defense, cell wall integrity, detoxification, JA-mediated signaling/defense, amino acid/carbohydrate metabolism, cell cycle regulation, and oxidative stress response. Recently, Chen et al. (2017) de novo assembled the pokeweed transcriptome from plants treated with or without cadmium. Although experimental variability was not addressed given the absence of RNA-seq replicates, authors reported DEGs and their pathway assignments. Of all pathways, photosynthesis was most enriched, with all 24 associated DEGs downregulated by cadmium treatment. These results reinforce the hypothesis that regulation of growth/defense balance is important for pokeweed stress tolerance.

### 5.3 Clarification of PAP isoforms and the broader role of PAP in pokeweed

In Neller et al. (2016), six de novo assembled transcripts were annotated as RIPs through BLASTx and protein domain searches. PAP-I and PAP-II coding regions were correct and complete, while partial transcripts of PAP-α and PAP-S were obtained. We reported a potentially novel PAP isoform, a transcript with 40% identity to PAP-α and 86% coverage, which we successfully cloned from pokeweed cDNA. We noted that another transcript, with 81% identity to PAP-I and 38% coverage, could represent an assembly artifact. Our results disagreed with previous reports of two PAP seed isoforms, and we did not identify PAP-R or PAP-III. However, we acknowledged that these discrepancies could arise from limiting our study to leaf tissue of young plants.

Through de novo assembly of the pokeweed genome in Neller et al. (2019), we further resolved the annotation of PAP isoforms. The genome was annotated using mixed-tissue pokeweed RNA-seq data (assembled via de novo and genome-guided approaches) and proteomes of RIP-containing Caryophylalles species. We reported genes of PAP-I, PAP-II, PAP-S1, PAP-S2, PAP-α, and a gene encoding the putative novel isoform described in Neller et al (2016). Evidence for PAP-H was seen in the edge assembly, but we found no evidence of additional isoforms, namely PAP-C, PAP-R, or PAP-III. This prompted a detailed literature review, which revealed that PAP-C, PAP-R, and PAP-I had identical N-terminal sequences and highly similar biochemical properties. Therefore, we suggested that reports of PAP-C and PAP-R referred to the same isoform. Curiously, the reported amino acid sequence of PAP-III was identical to that of PAP-II. Considering that PAP-II and PAP-III have different levels of antiviral activity and can be separated through chromatography (Rajamohan et al., 1999), we suggested that they represented post-translational modifications of the same isoform.

Moving from transcriptome- to genome-level analysis afforded greater insight into PAP seed isoforms. While we reported only PAP-S in Neller et al. (2016), our results supported two distinct seed isoforms in Neller et al. (2019). A gene with 77% identity to its best hit (PAP-I) was indeed PAP-S2, having 99% identity to the genomic clone sequence from Honjo et al. (2002). This gene is distinct from the presumed assembly artifact described in Neller et al. (2016), which has only 75% identity to published PAP-S2 (data not shown). We also observed that the seed isoforms had distinct responses to JA: PAP-S1 was most JA-induced of all isoforms, as observed previously, while PAP-S2, unlike all other PAP isoforms, was downregulated with JA. The seed isoforms were validated to be distinct through sequenced PCR products from gDNA and cDNA, but limitations of qPCR primer design allowed validation of differential expression only for PAP-S1. Therefore, RNA-seq analysis in Neller et al. (2019) appears to have teased apart expression dynamics undetectable through standard wet-lab methods. However, the situation is likely more complicated than we currently appreciate, given that PAP-S2 has two different transcription start sites, which we hypothesized in Neller et al (2019) to arise from alternate promoters.

In Neller et al. (2019), we gained insight into the relevance of distinct PAP isoforms by clustering the top DEGs based on their gene expression profiles. Two PAP-containing clusters were significantly enriched in GO terms. These clusters contained PAP-S2 and PAP-S1/PAP-α, which were associated with carbohydrate and amino acid metabolism, and JA signalling and terpenoid biosynthesis, respectively. The PAP-II cluster was most enriched in genes related to wounding response and JA biosynthesis. Curiously, the PAP-I cluster contained only nine genes, many of which were previously associated with plant growth responses. In Neller et al. (2019), we proposed that regulation of PAP expression in pokeweed is tied to the broader metabolic state of the plant. This regulation may be rooted in the shikimate pathway, which produces aromatic amino

acids from glycolysis intermediates, the former serving as precursors for important growth and defense compounds. We hypothesized that PAP, through its RIP activity, mediates a trade-off between plant growth and defense in pokeweed.

Further support for a link between PAP, JA signalling, and plant growth comes from previous studies in which non-toxic PAP mutants were stably expressed in non-heterologous plants. A PAP-I mutant (PAPn) lacking the ability to bind ribosomes and depurinate rRNA, but which maintained an intact active site, conferred resistance to viral and fungal infection in tobacco (Zoubenko et al., 2000). Authors referred to this resistance as a 'novel salicylic acid-independent pathway', associated with increased levels of basic PR proteins and wound-inducible protein kinase in PAPn-expressing plants. Although JA levels were not measured, these are characteristic markers of an active JA response. Further evidence comes from a report of transgenic Arabidopsis plants expressing a different PAP-I mutant (PAPc) that is capable of binding ribosomes but not depurinating them (Di and Tumer, 2015). Based on microarray analysis, expression of the JA biosynthesis gene *LOX3* was 4.6-fold higher in PAPc-expressing plants relative to wildtype Arabidopsis. The most upregulated genes in PAPc plants (4.9 – 6.1-fold) were auxin-responsive, with roles in auxin transport or auxin-induced transcription. Auxin, an indole derivative, is a classic plant growth hormone that coordinates tissue growth and organ development (Mroue et al., 2018). Recently, JA-mediated wound signalling was found to promote plant regeneration by increasing expression of an auxin biosynthesis gene (Zhang et al., 2019). In Neller et al. (2019), we found that the PAP-S2 cluster was significantly enriched in genes involved in indole transport and metabolism. Not only do these transgenic studies link PAP to JA signalling and plant growth, the use of PAP mutants reveals that this link is independent of ribosome depurination. Therefore, PAP-associated pathway activation is likely related to depurination of cellular mRNA, as PAP-I can

depurinate some capped RNAs (Hudak et al., 2000, 2002; Vivanco and Tumer, 2003). Curiously, transgenic tobacco expressing an active site mutant of PAP-I (PAPx, i.e. no depurination activity) also had enhanced levels of basic PRs (Zoubenko et al., 2000). This raises the intriguing possibility that PAP has an additional activity or is interacting with other cellular factors.

It is clear that RIP isoforms in several plants have discrete temporal, spatial, and stress-induced expression profiles (De Benito et al., 1998; Iglesias et al., 2008; Jiang et al., 2008; Kawade and Masuda, 2009; Parente et al., 2008; Qin et al., 2010; Tartarini et al., 2010). The finding that some RIPs are developmentally regulated and not stress-responsive supports the hypothesis of a role in plant growth. For example, in *Jatropha curcas*, Curcin is endosperm-specific and regulated temporally, while curcin-L is leaf-specific and induced by abiotic and biotic stresses; the antifungal activity of curcin-L is also higher than that of curcin (Qin et al., 2010). This mirrors the case of RIPs in spinach embryos, with transcriptional profiles of SoRIP1 and SoRIP2 consistent with discrete roles in embryo development and defense (Kawade and Masuda, 2009). We showed in Neller et al. (2019) that PAP isoforms had different responses to abiotic and biotic stresses and were co-expressed with different genes. This work complements reported differences in their temporal and tissue-related expression, summarized by Domashevskiy and Goss (2015), and in their activities on different RNA templates (Honjo et al., 2002; Rajamohan et al., 1999). Our identification of genes co-expressed with PAP isoforms provides a foundation to explore their unique roles in pokeweed.

### 5.4 Insight into PAP gene expression regulation

Heterologous expression of RIPs has been explored as a method to improve disease resistance in plants. Transgenic tobacco plants expressing PAP-I were resistant to viral and fungal pathogens; however, these plants displayed toxic phenotypes, such as stunted growth, mottled leaves, and sterility, correlated with the level of PAP synthesis (Lodge et al. 1993). Authors observed correct apoplastic localization of PAP in these plants, even though PAP was over-expressed using a strong, constitutive promoter. The issue of host toxicity has been overcome in some species using PAP mutants with reduced depurination ability (Dai et al., 2003; Zoubenko et al., 1997, 2000) or PAP isoforms that are less enzymatically active (Wang et al., 1998). However, these solutions do not utilize PAP to its full potential. More importantly, since unregulated toxicity of PAP is not observed in pokeweed, these results suggest that heterologous systems lack regulatory mechanisms to control PAP expression. Our work has provided insight into the regulation of PAP expression in its native plant.

A high level of RIP expression is presumably beneficial for the plant at certain times, based on their induced expression upon abiotic and biotic stress. In Neller et al. (2019), we identified stress-associated cis-regulatory elements (CREs) in the promoters of all PAP isoforms, as well as a promoter region of PAP-I required for induction by JA in tobacco. Therefore, regulation of PAP expression in pokeweed likely occurs at the transcriptional level via binding of stress-induced transcription factors. RIPs appear to be transcriptionally regulated in other plants, based on the requirement of certain promoter regions for stress-induced expression (Qin et al., 2009) and correlation of CREs with discrete isoform expression profiles (Kawade and Masuda, 2009). We also reported that PAP isoform gene models contained a long intron in the 5’ UTR, the presence of which enhanced PAP-I promoter-reporter expression in tobacco (Neller et al., 2019). Current

work in our lab is exploring the mechanism by which this enhancement occurs. Leader introns may increase RNA levels by promoting transcription initiation or increase protein levels by affecting mRNA processing, export, or stability (Gallegos and Rose, 2015; Laxa, 2017). A leader intron can also act as a classical enhancer (Kim et al., 2006) or alternate promoter (Morello et al., 2002, 2006; Qi et al., 2007). Based on RNA-seq quantification at both the transcript level (Neller et al., 2016) and gene level (Neller et al., 2019), we can conclude that PAP mRNA abundance increases in response to pathogen-associated stress. This JA induction of PAP mRNA correlates with an increase in PAP at the protein level (Klenov et al., 2015).

Owing to the toxicity of RIPs, we would expect that certain mechanisms exist to limit their expression or enzymatic activity. Ideally, these would prevent RIP function when it is not required and enable a finely tuned response upon induction. Significant evidence has accumulated for the inhibition of RIPs at the protein level. RIPs are often sequestered from host ribosomes, demonstrated by their extracellular or vacuolar localization (Carzaniga et al., 1994; Di Cola et al., 1999; Ready et al., 1986; Yoshinari et al., 2009). RIPs in cereal plants are generally cytosolic; therefore, they require processing prior to activation and tend to be inactive on conspecific ribosomes (de Zaeytijd and Van Damme, 2017). In pokeweed, PAP dimerization in the cytosol is thought to limit depurination of pokeweed ribosomes that could arise from retrotranslocation (Tourlakis et al., 2010). However, little is known about how PAP expression, or that of RIPs in general, is fine-tuned in a plant. In yeast, over-expression of PAP, but not an enzymatically inactive mutant, specifically reduced the level of PAP mRNA (Parikh et al., 2002). Therefore, PAP may depurinate its own mRNA in pokeweed. Preliminary evidence supports the possibility of RIP auto-depurination in a native plant: ME1 purified from *Mirabilis expansa* depurinated ME1 mRNA in vitro (Vepachedu et al., 2003). Authors hypothesized that auto-depurination could explain the

curious observation of JA-induced expression of ME1 RNA without a corresponding increase in ME1 protein. That is, perhaps the transcript was present but unable to be translated because of depurinated nucleotides. A discrepancy between the endogenous levels of RIP transcript and protein was also observed in sugar beet (Iglesias et al., 2008). Beetin transcripts were present at similar levels throughout plant development, but the protein was only synthesized in adult plants. Furthermore, stress-induced expression of beetin, at both mRNA and protein levels, occurred only in adult plants, confirming a developmentally controlled, post-transcriptional regulatory mechanism.

Our work has provided insight into mechanisms by which PAP expression may be regulated in pokeweed. Consistent with previous reports of post-transcriptional regulation of RIPs, we demonstrated that PAP-I mRNA undergoes small RNA-mediated transcript cleavage in pokeweed (Klenov et al., 2015). Since levels of the small RNA and PAP-I mRNA were positively correlated in response to JA, this could represent a mechanism to limit the extent to which PAP-I mRNA abundance can increase from transcription. This may reflect a broader regulatory dynamic to limit induced expression of defense genes in pokeweed, as we found that approximately half of JA-responsive miRNA/target pairs had positively correlated expression (Neller et al., 2018). Additional findings at the transcriptome and genomic levels suggest further opportunities for PAP expression regulation. In Neller et al. (2016), we reported NATs of PAP isoforms. As seen with the levels of PAP-I mRNA/small RNA, expression of each isoform and its corresponding NAT was positively correlated in response to JA. This introduces the possibility of a novel PAP regulatory mechanism, given that NATs can modulate gene expression at various levels, including transcription, small RNA interference, mRNA splicing, or RNA stability (Liu et al., 2015; Rosikiewicz and Makałowska, 2016; Zhang et al., 2013b). In Neller et al. (2019), we identified

PAP pseudogenes that were transcriptionally active and stress-responsive despite encoding numerous stop codons. One pseudogene was assigned to the same cluster as a PAP isoform, i.e. they were co-expressed. Like NATs, pseudogenes can regulate expression of functional parental genes through various mechanisms (Kovalenko and Patrushev, 2018). Both pseudogene transcripts and NATs can generate double-stranded RNA through hybridization with the parental transcript, which can lead to siRNAs that target the parental gene (Kovalenko and Patrushev, 2018; Zhang et al., 2013b). Pseudogenes that produce a sense-oriented transcript can also act as a small RNA 'sponge' for the parental transcript (An et al., 2017). Given that we did not identify a miRNA targeting PAP-I (Neller et al., 2018), the small RNA may have siRNA origins, perhaps derived from a PAP NAT or pseudogene. Taken together, results from this project have elucidated novel mechanisms by which PAP gene expression may be regulated in pokeweed, and these may extend to other RIP-synthesizing plants.

### 5.5 Future Directions

We investigated transcript-level and gene-level differential expression in Neller et al. (2016) and (2019), respectively. Using our current data, a logical extension to this analysis would be an investigation of transcriptome complexity arising from processes such as alternative transcription and alternative splicing. Recent studies have applied RNA-seq to investigate this subject in plants (Song et al., 2019). Using computational tools, events arising from these biological processes can be identified and classified. Subsequently, differential expression analysis can be performed to identify cases of isoform switching. To bypass the error-prone process of transcript assembly, recent approaches have adopted the use of long-read technology (An et al., 2018). Long reads or mate-pair reads could also be applied to increase the contiguity of the current

pokeweed genome assembly. In addition to improving its utility as a reference genome, this would enable wider comparative studies across plants. One intriguing question is whether RIP-containing gene clusters are conserved, and this could be addressed by investigating syntenic regions across RIP-synthesizing plants.

In Neller et al. (2019), we showed that PAP isoforms had distinct gene expression profiles in response to stress treatments and were co-expressed with different genes. To gain more insight into the relevance of PAP isoforms in pokeweed, this analysis could be repeated with additional stress treatments, tissues, and developmental time-points. Importantly, we have yet to identify a condition under which the novel PAP isoform is induced, and its promoter is currently unknown due to fragmentation of the genome assembly. To clarify the relevance of PAP isoforms, they can be independently expressed in tobacco or Arabidopsis. Although this was conducted previously for PAP-I and PAP-II, expression was under control of the 35S CaMV promoter (Lodge et al., 1993; Wang et al., 1998). With availability of the endogenous promoter of each isoform, their expression profiles could be determined in-planta via stable integration of promoter-reporter constructs. Using a reporter gene would avoid previously encountered issues of host toxicity due to RIP activity. Ideally, PAP isoforms in pokeweed would be knocked down/out via RNA silencing or CRISPR-based gene-editing. However, it may be difficult to achieve knockout of all isoforms, and even more challenging to independently assess isoforms.

Previous studies revealed that transgenic plants expressing PAP-I mutants unable to depurinate rRNA, but maintaining intact active sites, had upregulation of genes involved in pathogen defense and plant growth (Di and Tumer, 2015; Zoubenko et al., 2000). This observation is likely a result of PAP depurinating cellular mRNA transcripts, given its ability to depurinate capped and uncapped viral RNA (Hudak et al., 2002; Wang and Hudak, 2006), and probably its

own transcript (Parikh et al., 2002). Our group has proposed a high-throughput approach to detect transcriptome-wide depurination events (Jobst et al., 2016). Using this approach, transcripts depurinated in PAP-expressing cells could be identified relative to control cells expressing PAPx, an enzymatically inactive mutant. This experiment could be performed in tandem with active ribosome profiling (Clamer et al., 2018) to assess whether PAP-induced depurination affects the pool of translated transcripts. Unlike traditional ribosome profiling (Ingolia et al., 2009), active profiling would distinguish translating ribosomes from stalled ribosomes, the latter a consequence of mRNA depurination (Jobst et al., 2016). The functional relevance of PAP-targeted transcripts could be assessed through GO enrichment analysis. Of particular interest would be the effect of JA on transcript targeting by PAP, and how this influences the pool of translated mRNAs. I hypothesize that in response to JA, transcripts targeted by PAP would be enriched in GO terms associated with growth/housekeeping roles. Conversely, I hypothesize that transcripts undergoing JA-induced translation would be enriched in defense responses, and that this translation-level effect requires PAP-induced depurination.

**References**

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscisic Acid Signaling. *Plant Cell* 15, 63-78. doi:10.1105/tpc.006130.

Abuqamar, S., Chen, X., Dhawan, R., Bluhm, B., Salmeron, J., Lam, S., et al. (2006). Expression profiling and mutant analysis reveals complex regulatory networks involved in Arabidopsis response to Botrytis infection. *Plant J.* 48, 28–44. doi:10.1111/j.1365-313X.2006.02849.x.

Agarwal, M., Hao, Y., Kapoor, A., Dong, C. H., Fujii, H., Zheng, X., et al. (2006). A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J. Biol. Chem.* 281, 37636-45. doi:10.1074/jbc.M605895200.

Akua, T., Berezin, I., and Shaul, O. (2010). The leader intron of AtMHX can elicit, in the absence of splicing, low-level intron-mediated enhancement that depends on the internal intron sequence. *BMC Plant Biol.* 10. doi:10.1186/1471-2229-10-93.

Ali, S., Ganai, B. A., Kamili, A. N., Bhat, A. A., Mir, Z. A., Bhat, J. A., et al. (2018). Pathogenesis-related proteins and peptides as promising tools for engineering plants with multiple stress tolerance. *Microbiol. Res.* 212–213, 29–37. doi:10.1016/j.micres.2018.04.008.

Alison Dunn, M., White, A. J., Vural, S., and Hughes, M. A. (1998). Identification of promoter elements in a low-temperature-responsive gene (blt4.9) from barley (Hordeum vulgare L.). *Plant Mol. Biol.* 38, 551-64. doi:10.1023/A:1006098132352.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi:10.1093/nar/25.17.3389.

An, D., Cao, H., Li, C., Humbeck, K., and Wang, W. (2018). Isoform Sequencing and State-of-Art Applications for Unravelling Complexity of Plant Transcriptomes. *Genes (Basel).* 9, 43. doi:10.3390/genes9010043.

An, Y., Furber, K. L., and Ji, S. (2017). Pseudogenes regulate parental gene expression via ceRNA network. *J. Cell. Mol. Med.* 21, 185–192. doi:10.1111/jcmm.12952.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi:10.1093/bioinformatics/btu638.

Anderson, J. P., Badruzsaufari, E., Schenk, P. M., Manners, J. M., Desmond, O. J., Ehlert, C., et al. (2004). Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. *Plant Cell* 16, 3460–3479. doi:10.1105/tpc.104.025833.

Aoto, Y., Hachiya, T., Okumura, K., Hase, S., Sato, K., Wakabayashi, Y., et al. (2017). DEclust: A statistical approach for obtaining differential expression profiles of multiple conditions. *PLoS One* 12. doi:10.1371/journal.pone.0188285.

Aoyama, T., Dong, C.-H., Wu, Y., Carabelli, M., Sessa, G., Ruberti, I., et al. (2007). Ectopic Expression of the Arabidopsis Transcriptional Activator Athb-1 Alters Leaf Cell Fate in Tobacco. *Plant Cell* 7, 1773. doi:10.2307/3870186.

Asha, S., Sreekumar, S., and Soniya, E. V (2016). Unravelling the complexity of microRNA-mediated gene regulation in black pepper (Piper nigrum L.) using high-throughput small RNA profiling. *Plant Cell Rep.* 35, 53–63. doi:10.1007/s00299-015-1866-x.

Audenaert, K., De Meyer, G. B., and Höfte, M. M. (2002). Abscisic acid determines basal susceptibility of tomato to Botrytis cinerea and suppresses salicylic acid-dependent signaling mechanisms. *Plant Physiol.* 128, 491–501. doi:10.1104/pp.010605.

Axtell, M. J. (2013). Classification and Comparison of Small RNAs from Plants. *Annu. Rev. Plant Biol.* 64, 137-59. doi:10.1146/annurev-arplant-050312-120043.

Ayashi, T. H., Obayashi, D. K., Ariu, T. K., Ahara, M. T., Ada, K. H., Ouzuma, Y. K., et al. (2003). Genomic Cloning of Ribonucleases in Nicotiana glutinosa Leaves, as Induced in Response to Wounding or to TMV-Infection, and Characterization of Their Promoters. *Biosci. Biotechnol. Biochem.* 67, 2574–2583. doi:10.1271/bbb.67.2574.

Backer, R., Naidoo, S., and van den Berg, N. (2019). The NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1 (NPR1) and Related Family: Mechanistic Insights in Plant Disease Resistance. *Front. Plant Sci.* 10. doi:10.3389/fpls.2019.00102.

Bairoch, A., and Boeckmann, B. (1991). The Swiss-Prot Protein-Sequence Data-Bank. *Nucleic Acids Res.* 22, 3093–3096.

Bakan, B., Hamberg, M., Perrocheau, L., Maume, D., Rogniaux, H., Tranquet, O., et al. (2006). Specific adduction of plant lipid transfer protein by an allene oxide generated by 9-lipoxygenase and allene oxide synthase. *J. Biol. Chem.* 281, 38981–38988. doi:10.1074/jbc.M608580200.

Baldrich, P., Beric, A., and Meyers, B. C. (2018). Despacito: the slow evolutionary changes in plant microRNAs. *Curr. Opin. Plant Biol.* 42, 16–22. doi:10.1016/j.pbi.2018.01.007.

Barbieri, L., Aron, G. M., Irvin, J. D., and Stirpe, F. (1982). Purification and partial characterization of another form of the antiviral protein from the seeds of Phytolacca americana L. (pokeweed). *Biochem. J.* 203, 55–59.

Barbieri, L., Bolognesi, A., Cenini, P., Falasca, A. I., Minghetti, A., Garofano, L., et al. (1989). Ribosome-inactivating proteins from plant cells in culture. *Biochem. J.* 257, 801–807. doi:10.1042/bj2570801.

Bekeredjian-Ding, I., Foermer, S., Kirschning, C. J., Parcina, M., and Heeg, K. (2012). Poke weed mitogen requires toll-like receptor ligands for proliferative activity in human and murine b lymphocytes. *PLoS One* 7. doi:10.1371/journal.pone.0029806.

Bennett, M. (2000). Nuclear DNA Amounts in Angiosperms and their Modern Uses, 807 New Estimates. *Ann. Bot.* 86, 859–909. doi:10.1006/anbo.2000.1253.

Blankenberg, D., Kuster, G. Von, Coraor, N., Ananda, G., Lazarus, R., Mangan, M., et al. (2010). Galaxy: A web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol.*

*Biol.* 19, 1-21. doi:10.1002/0471142727.mb1910s89.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170.

Bolger, M. E., Arsova, B., and Usadel, B. (2018). Plant genome and transcriptome annotations: From misconceptions to simple solutions. *Brief. Bioinform.* 19, 437–449. doi:10.1093/bib/bbw135.

Bologna, N. G., Iselin, R., Abriata, L. A., Sarazin, A., Pumplin, N., Jay, F., et al. (2018). Nucleo-cytosolic Shuttling of ARGONAUTE1 Prompts a Revised Model of the Plant MicroRNA Pathway. *Mol. Cell* 69, 709–719.e5. doi:10.1016/j.molcel.2018.01.007.

Bolognesi, A., Barbieri, L., Abbondanza, A., Falasca, A. I., Carnicelli, D., Battelli, M. G., et al. (1990). Purification and properties of new ribosome-inactivating proteins with RNA N-glycosidase activity. *Biochim. Biophys. Acta* 1087, 293–302. doi:10.1016/0167-4781(90)90002-J.

Bonness, M. S., Ready, M. P., Irvin, J. D., and Mabry, T. J. (1994). Pokeweed antiviral protein inactivates pokeweed ribosomes; implications for the antiviral mechanism. *Plant J.* 5, 173–183. doi:10.1046/j.1365-313X.1994.05020173.x.

Boter, M., Ruíz-Rivero, O., Abdeen, A., and Prat, S. (2004). Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. *Genes Dev.* 18, 1577–1591. doi:10.1101/gad.297704.

Bozorov, T. A., Baldwin, I. T., and Kim, S. G. (2012). Identification and profiling of miRNAs during herbivory reveals jasmonate-dependent and -independent patterns of accumulation in Nicotiana attenuata. *BMC Plant Biol.* 12, 209. doi:10.1186/1471-2229-12-209.

Buchel, A. S., Brederode, F. T., Bol, J. F., and Linthorst, H. J. M. (1999). Mutation of GT-1 binding sites in the Pr-1a promoter influences the level of inducible gene expression in vivo. *Plant Mol. Biol.* 40:387-96. doi:10.1023/A:1006144505121.

Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., et al. (2016). JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.* 17. doi:10.1186/s13059-016-0924-1.

Campbell, M. S., Holt, C., Moore, B., and Yandell, M. (2014a). Genome Annotation and Curation Using MAKER and MAKER-P. *Curr. Protoc. Bioinforma.* 2014, 4.11.1-4.11.39. doi:10.1002/0471250953.bi0411s48.

Campbell, M. S., Law, M., Holt, C., Stein, J. C., Moghe, G. D., Hufnagel, D. E., et al. (2014b). MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* 164, 513–24. doi:10.1104/pp.113.230144.

Carr, J. P., Murphy, A. M., Tungadi, T., and Yoon, J. Y. (2019). Plant defense signals: Players and pawns in plant-virus-vector interactions. *Plant Sci.* 279, 87–95. doi:10.1016/j.plantsci.2018.04.011.

Carzaniga, R., Sinclair, L., Fordham-Skelton, A. P., Harris, N., and Croy, R. R. D. (1994). Cellular and subcellular distribution of saporins, type-1 ribosome-inactivating proteins, in

soapwort (Saponaria officinalis L.). *Planta* 194, 461–470. doi:10.1007/BF00714457.

Cercós, M., Gómez-Cadenas, A., and Ho, T. H. D. (1999). Hormonal regulation of a cysteine proteinase gene, EPB-1, in barley aleurone layers: Cis- and trans-acting elements involved in the co-ordinated gene expression regulated by gibberellins and abscisic acid. *Plant J.* 19, 107–118. doi:10.1046/j.1365-313X.1999.00499.x.

Chakravarthy, S., Tuori, R., D'Ascenzo, M., Fobert, P., Despres, C., and Martin, G. (2003). The Tomato Transcription Factor Pti4 Regulates Defense-Related Gene Expression via GCC Box and Non-GCC Box cis Elements. *Plant Cell*. doi:10.1105/tpc.017574.

Champigny, M. J., Isaacs, M., Carella, P., Faubert, J., Fobert, P. R., and Cameron, R. K. (2013). Long distance movement of DIR1 and investigation of the role of DIR1-like during systemic acquired resistance in Arabidopsis. *Front. Plant Sci.* 4. doi:10.3389/fpls.2013.00230.

Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., et al. (2010). Draft genome sequence of the oilseed species Ricinus communis. *Nat. Biotechnol.* 28, 951–6. doi:10.1038/nbt.1674.

Chen, X. (2004). A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. *Science* 303, 2022–2025. doi:10.1126/science.1088060.

Chen, Y., Zhi, J., Zhang, H., Li, J., Zhao, Q., and Xu, J. (2017). Transcriptome analysis of Phytolacca americana L. in response to cadmium stress. *PLoS One* 12. doi:10.1371/journal.pone.0184681.

Chiba, Y., and Shimizu, T. (2015). Identification of Arabidopsis thaliana NRT1 / PTR FAMILY ( NPF ) proteins capable of transporting plant hormones. *J. Plant Res.*, 679–686. doi:10.1007/s10265-015-0710-2.

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., et al. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* 448, 666–671. doi:10.1038/nature06006.

Chini, A., Gimenez-Ibanez, S., Goossens, A., and Solano, R. (2016). Redundancy and specificity in jasmonate signalling. *Curr. Opin. Plant Biol.* 33, 147–156. doi:10.1016/j.pbi.2016.07.005.

Chini, A., Monte, I., Zamarreño, A. M., Hamberg, M., Lassueur, S., Reymond, P., et al. (2018). An OPR3-independent pathway uses 4,5-didehydrojasmonate for jasmonate synthesis. *Nat. Chem. Biol.* 14, 171–178. doi:10.1038/nchembio.2540.

Chuang, W.-P., Herde, M., Ray, S., Castano-Duque, L., Howe, G. A, and Luthe, D. S. (2014). Caterpillar attack triggers accumulation of the toxic maize protein RIP2. *New Phytol.* 201, 928–39. doi:10.1111/nph.12581.

Clamer, M., Tebaldi, T., Lauria, F., Bernabò, P., Gómez-Biagi, R. F., Marchioretto, M., et al. (2018). Active Ribosome Profiling with RiboLace. *Cell Rep.* 25, 1097–1108.e5. doi:10.1016/j.celrep.2018.09.084.

Clouse, J. W., Adhikary, D., Page, J. T., Ramaraj, T., Deyholos, M. K., Udall, J. A., et al. (2016).

The Amaranth Genome: Genome, Transcriptome, and Physical Map Assembly. *Plant Genome* 9, 0. doi:10.3835/plantgenome2015.07.0062.

Coleman, G. D., Chen, T. H., Ernst, S. G., and Fuchigami, L. (1991). Photoperiod control of poplar bark storage protein accumulation. *Plant Physiol.* 96, 686–692. doi:10.1104/pp.96.3.686.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi:10.1093/bioinformatics/bti610.

Côté, C., and Rutledge, R. G. (2003). An improved MUG fluorescent assay for the determination of GUS activity within transgenic tissue of woody plants. *Plant Cell Rep.* 21, 619–24. doi:10.1007/s00299-002-0543-z.

Dai, W. D., Bonos, S., Guo, Z., Meyer, W. A., Day, P. R., and Belanger, F. C. (2003). Expression of pokeweed antiviral proteins in creeping bentgrass. *Plant Cell Rep.* 21, 497–502. doi:10.1007/s00299-002-0534-0.

Dai, X., and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, W155-9. doi:10.1093/nar/gkr319.

De Benito, F. M., Iglesias, R., Ferreras, J. M., Citores, L., Camafeita, E., Méndez, E., et al. (1998). Constitutive and inducible type 1 ribosome-inactivating proteins (RIPs) in elderberry (Sambucus nigra L.). *FEBS Lett.* 428, 75–79. doi:10.1016/S0014-5793(98)00496-7.

de Zaeytijd, J., and Van Damme, E. J. M. (2017). Extensive evolution of cereal ribosome-inactivating proteins translates into unique structural features, activation mechanisms, and physiological roles. *Toxins (Basel).* 9. doi:10.3390/toxins9040123.

Dechorgnat, J., Chardon, F., and Gaufichon, L. (2010). Nitrogen uptake , assimilation and remobilization in plants : challenges for sustainable and productive agriculture. *Ann bot.* 105, 1141–1157. doi:10.1093/aob/mcq028.

Dempsey, D. A., Vlot, A. C., Wildermuth, M. C., and Klessig, D. F. (2012). Salicylic Acid Biosynthesis and Metabolism. *Arab. B.* 9, e0156. doi:10.1199/tab.0156.

Denancé, N., Ranocha, P., Oria, N., Barlet, X., Rivière, M. P., Yadeta, K. A., et al. (2013). Arabidopsis wat1 (walls are thin1)-mediated resistance to the bacterial vascular pathogen, Ralstonia solanacearum, is accompanied by cross-regulation of salicylic acid and tryptophan metabolism. *Plant J.* 73, 225–239. doi:10.1111/tpj.12027.

Derksen, H., Rampitsch, C., and Daayf, F. (2013). Signaling cross-talk in plant disease resistance. *Plant Sci.* 207, 79–87. doi:10.1016/j.plantsci.2013.03.004.

Despres, C., DeLong, C., Glaze, S., Liu, E., and Fobert, P. R. (2007). The Arabidopsis NPR1/NIM1 Protein Enhances the DNA Binding Activity of a Subgroup of the TGA Family of bZIP Transcription Factors. *Plant Cell* 12, 279. doi:10.2307/3870928.

Di, R., and Tumer, N. E. (2015). Pokeweed antiviral protein: Its cytotoxicity mechanism and applications in plant disease resistance. *Toxins (Basel).* 7, 755–772.

doi:10.3390/toxins7030755.

Di Cola, A., Marcozzi, G., Balestrini, R., and Spanò, L. (1999). Localization of the type I ribosome-inactivating protein, luffin, in adult and embryonic tissues of Luffa cylindrica L. Roem. *J. Exp. Bot.* 50, 573–579. doi:10.1093/jxb/50.334.573.

Di Maro, A., Citores, L., Russo, R., Iglesias, R., and Ferreras, J. M. (2014). Sequence comparison and phylogenetic analysis by the Maximum Likelihood method of ribosome-inactivating proteins from angiosperms. *Plant Mol. Biol.* 85, 575–88. doi:10.1007/s11103-014-0204-y.

Ding, Y., Sun, T., Ao, K., Peng, Y., Zhang, Y., Li, X., et al. (2018). Opposite Roles of Salicylic Acid Receptors NPR1 and NPR3/NPR4 in Transcriptional Regulation of Plant Immunity. *Cell* 173, 1454–1467.e10. doi:10.1016/j.cell.2018.03.044.

Dohm, J. C., Minoche, A. E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., et al. (2014). The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature* 505, 546–9. doi:10.1038/nature12817.

Dolezel, J., Binarova, P., and Lucretti, S. (1989). Analysis of Nuclear DNA Content in Plant Cells by Flow Cytometry. *Biol. Plant.* 31, 113–120. doi:10.1007/BF02907241.

Domashevskiy, A. V., and Goss, D. J. (2015). Pokeweed antiviral protein, a ribosome inactivating protein: Activity, inhibition and prospects. *Toxins (Basel).* 7, 274–298. doi:10.3390/toxins7020274.

Dong, J., Chen, C., and Chen, Z. (2003). Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Mol. Biol.* 51, 21–37. doi:10.1023/A:1020780022549.

Dou, C. M., Fu, X. P., Chen, X. C., Shi, J. Y., and Chen, Y. X. (2009). Accumulation and detoxification of manganese in hyperaccumulator Phytolacca americana. *Plant Biol. (Stuttg).* 11, 664–670. doi:10.1111/j.1438-8677.2008.00163.x.

Duan, C.-G., Wang, X., Zhang, L., Xiong, X., Zhang, Z., Tang, K., et al. (2017). A protein complex regulates RNA processing of intronic heterochromatin-containing genes in Arabidopsis . *Proc. Natl. Acad. Sci.* 114, E7377–E7384. doi:10.1073/pnas.1710683114.

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al. (2003). OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* doi:10.1046/j.1365-313X.2003.01661.x.

Dunaeva, M., Goebel, C., Wasternack, C., Parthier, B., and Goerschen, E. (1999). The jasmonate-induced 60 kDa protein of barley exhibits N-glycosidase activity in vivo. *FEBS Lett.* 452, 263–6.

Ellis, C. M., Nagpal, P., Young, J. C., Hagen, G., Guilfoyle, T. J., and Reed, J. W. (2005). AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission in Arabidopsis thaliana. *Development* 132, 4563–4574. doi:10.1242/dev.02012.

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16. doi:10.1186/s13059-015-0721-2.

Endo, Y., Tsurugi, K., and Lambert, J. M. (1988). The site of action of six different ribosome-inactivating proteins from plants on eukaryotic ribosomes: the RNA N-glycosidase activity of the proteins. *Biochem. Biophys. Res. Commun.* 150, 1032–1036. doi:10.1016/0006-291X(88)90733-4.

Erb, M., Veyrat, N., Robert, C. A. M., Xu, H., Frey, M., Ton, J., et al. (2015). Indole is an essential herbivore-induced volatile priming signal in maize. *Nat. Commun.* 6. doi:10.1038/ncomms7273.

Eulgem, T., Tsuchiya, T., Wang, X. J., Beasley, B., Cuzick, A., Tör, M., et al. (2007). EDM2 is required for RPP7-dependent disease resistance in Arabidopsis and affects RPP7 transcript levels. *Plant J.* 49, 829–839. doi:10.1111/j.1365-313X.2006.02999.x.

Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K., and Rask, L. (1999). Interaction between composite elements in the napA promoter: Both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol. Biol.* doi:10.1023/A:1006206124512.

Farrell, J. D., Byrne, S., Paina, C., and Asp, T. (2014). De novo assembly of the perennial ryegrass transcriptome using an RNA-Seq strategy. *PLoS One* 9, e103567. doi:10.1371/journal.pone.0103567.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: The protein families database. *Nucleic Acids Res.* 42. doi:10.1093/nar/gkt1223.

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* 39. doi:10.1093/nar/gkr367.

Finotello, F., and Di Camillo, B. (2015). Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Brief. Funct. Genomics* 14, 130–42. doi:10.1093/bfgp/elu035.

Fonseca, S., Chini, A., Hamberg, M., Adie, B., Porzel, A., Kramell, R., et al. (2009). (+)-7-iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. *Nat. Chem. Biol.* 5, 344–350. doi:10.1038/nchembio.161.

Foroughi, S., Baker, A. J., Roessner, U., Johnson, A. A., Bacic, A., and Callahan, D. L. (2014). Hyperaccumulation of zinc by Noccaea caerulescens results in a cascade of stress responses and changes in the elemental profile. *Metallomics* 6, 1671–82. doi:10.1039/c4mt00132j.

Friedländer, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., et al. (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.* 26, 407–415. doi:10.1038/nbt1394.

Fu, Z. Q., Yan, S., Saleh, A., Wang, W., Ruble, J., Oka, N., et al. (2012). NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. *Nature*. doi:10.1038/nature11162.

Fugate, K. K., Fajardo, D., Schlautman, B., Ferrareze, J. P., Bolton, M. D., Campbell, L. G., et

al. (2014). Generation and Characterization of a Sugarbeet Transcriptome and Transcript-Based SSR Markers. *Plant Genome* 7. doi:10.3835/plantgenome2013.11.0038.

Fujii, H., Chinnusamy, V., Rodrigues, A., Rubio, S., Antoni, R., Park, S.-Y., et al. (2009). In vitro reconstitution of an abscisic acid signalling pathway. *Nature* 462, 660–664. doi:10.1038/nature08599.

Fujii, H., Chiou, T.-J., Lin, S.-I., Aung, K., and Zhu, J.-K. (2005). A miRNA involved in phosphate-starvation response in Arabidopsis. *Curr. Biol.* 15, 2038–43. doi:10.1016/j.cub.2005.10.016.

Fukudome, A., and Fukuhara, T. (2017). Plant dicer-like proteins: double-stranded RNA-cleaving enzymes for small RNA biogenesis. *J. Plant Res.* 130, 33–44. doi:10.1007/s10265-016-0877-1.

Furihata, T., Maruyama, K., Fujita, Y., Umezawa, T., Yoshida, R., Shinozaki, K., et al. (2006). Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1988–1993. doi:10.1073/pnas.0505667103.

Galiana, E., Bonnet, P., Conrod, S., Keller, H., Panabières, F., Ponchet, M., et al. (1997). RNase activity prevents the growth of a fungal pathogen in tobacco leaves and increases upon induction of systemic acquired resistance with elicitin. *Plant Physiol.* 115, 1557–1567. doi:10.1104/pp.115.4.1557.

Gallegos, J. E., and Rose, A. B. (2015). The enduring mystery of intron-mediated enhancement. *Plant Sci.* 237, 8–15. doi:10.1016/j.plantsci.2015.04.017.

Gandikota, M., Birkenbihl, R. P., Höhmann, S., Cardon, G. H., Saedler, H., and Huijser, P. (2007). The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. *Plant J.* 49, 683–93. doi:10.1111/j.1365-313X.2006.02983.x.

Gao, L., Peng, K., Chen, Y., Wang, G., and Shen, Z. (2012). Roles of apoplastic peroxidases, laccases, and lignification in the manganese tolerance of hyperaccumulator Phytolacca americana. *Acta Physiol. Plant.* 34, 151–159. doi:10.1007/s11738-011-0813-x.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* 15, 1451–1455. doi:10.1101/gr.4086505.

Girault, T., François, J., Rogniaux, H., Pascal, S., Delrot, S., Coutos-Thévenot, P., et al. (2008). Exogenous application of a lipid transfer protein-jasmonic acid complex induces protection of grapevine towards infection by Botrytis cinerea. *Plant Physiol. Biochem.* 46, 140–149. doi:10.1016/j.plaphy.2007.10.005.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11, R86. doi:10.1186/gb-2010-11-8-r86.

Gou, M., Su, N., Zheng, J., Huai, J., Wu, G., Zhao, J., et al. (2009). An F-box gene, CPR30, functions as a negative regulator of the defense response in Arabidopsis. *Plant J.* 60, 757–

770. doi:10.1111/j.1365-313X.2009.03995.x.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi:10.1038/nbt.1883.

Gu, Y. Q., Wildermuth, M. C., Chakravarthy, S., Loh, Y. T., Yang, C., He, X., et al. (2002). Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in Arabidopsis. *Plant Cell*. doi:10.1105/tpc.000794.

Gubler, F., Kalla, R., Roberts, J. K., and Jacobsen, J. V. (1995). Gibberellin-Regulated Expression of a myb Gene in Barley Aleurone Cells: Evidence for Myb Transactivation of a High-pl alpha-Amylase Gene Promoter. *Plant Cell*. doi:10.1105/tpc.7.11.1879.

Guzman, F., Almerão, M. P., Korbes, A. P., Christoff, A. P., Zanella, C. M., Bered, F., et al. (2013). Identification of potential miRNAs and their targets in Vriesea carinata (Poales, Bromeliaceae). *Plant Sci.* 210, 214–23. doi:10.1016/j.plantsci.2013.05.013.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–512. doi:10.1038/nprot.2013.084.

Hackenberg, M., Huang, P. J., Huang, C. Y., Shi, B. J., Gustafson, P., and Langridge, P. (2013). A Comprehensive Expression Profile of MicroRNAs and Other Classes of Non-Coding Small RNAs in Barley Under Phosphorous-Deficient and -Sufficient Conditions. *DNA Res.* 109–125.

Hamiaux, C., Drummond, R. S. M., Janssen, B. J., Ledger, S. E., Cooney, J. M., Newcomb, R. D., et al. (2012). DAD2 is an α/β hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. *Curr. Biol.* 22, 2032–2036. doi:10.1016/j.cub.2012.08.007.

Hanson, J., Hanssen, M., Wiese, A., Hendriks, M. M. W. B., and Smeekens, S. (2008). The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of ASPARAGINE SYNTHETASE1 and PROLINE DEHYDROGENASE2. *Plant J.* 53, 935–949. doi:10.1111/j.1365-313X.2007.03385.x.

He, Y.-W., Guo, C.-X., Pan, Y.-F., Peng, C., and Weng, Z.-H. (2008). Inhibition of hepatitis B virus replication by pokeweed antiviral protein in vitro. *World J. Gastroenterol.* 14, 1592–1597. doi:10.3748/wjg.14.1592.

Healey, A., Furtado, A., Cooper, T., and Henry, R. J. (2014). Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* 10. doi:10.1186/1746-4811-10-21.

Heim, M. A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P. C. (2003). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* 20, 735–747. doi:10.1093/molbev/msg088.

Hickman, R., van Verk, M. C., Van Dijken, A. J. H., Pereira Mendes, M., Vroegop-Vos, I. A., Caarls, L., et al. (2017). Architecture and Dynamics of the Jasmonic Acid Gene Regulatory

Network. *Plant Cell*, tpc.00958.2016. doi:10.1105/tpc.16.00958.

Hobo, T., Asada, M., Kowyama, Y., and Hattori, T. (1999). ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. *Plant J.* doi:10.1046/j.1365-313X.1999.00565.x.

Hogg, T., Mendel, J. T., and Lavezo, J. L. (2015). Structural analysis of a type 1 ribosome inactivating protein reveals multiple L-asparagine-N-acetyl-D-glucosamine monosaccharide modifications: Implications for cytotoxicity. *Mol. Med. Rep.* 12, 5737–5745. doi:10.3892/mmr.2015.4146.

Honjo, E., Dong, D., Motoshima, H., and Watanabe, K. (2002). Genomic clones encoding two isoforms of pokeweed antiviral protein in seeds (PAP-S1 and S2) and the N-glycosidase activities of their recombinant proteins on ribosomes and DNA in comparison with other isoforms. *J. Biochem*. 131, 225–231.

Hotto, A. M., Schmitz, R. J., Fei, Z., Ecker, J. R., and Stern, D. B. (2011). Unexpected Diversity of Chloroplast Noncoding RNAs as Revealed by Deep Sequencing of the Arabidopsis Transcriptome. *G3 (Bethesda).* 1, 559–570. doi:10.1534/g3.111.000752.

Hu, Q., Min, L., Yang, X., Jin, S., Zhang, L., Li, Y., et al. (2018). Laccase GhLac1 Modulates Broad-Spectrum Biotic Stress Tolerance via Manipulating Phenylpropanoid Pathway and Jasmonic Acid Synthesis. *Plant Physiol.* 176, 1808–1823. doi:10.1104/pp.17.01628.

Hudak, K. A., Bauman, J. D., and Tumer, N. E. (2002). Pokeweed antiviral protein binds to the cap structure of eukaryotic mRNA and depurinates the mRNA downstream of the cap. *RNA* 8, 1148–1159. doi:10.1017/S1355838202026638.

Hudak, K. A., Wang, P., and Tumer, N. E. (2000). A novel mechanism for inhibition of translation by pokeweed antiviral protein: Depurination of the capped RNA template. *RNA* 6, 369–380. doi:10.1017/S1355838200991337.

Iglesias, R., Pérez, Y., Citores, L., Ferreras, J. M., Méndez, E., and Girbés, T. (2008). Elicitor-dependent expression of the ribosome-inactivating protein beetin is developmentally regulated. *J. Exp. Bot.* 59, 1215–23. doi:10.1093/jxb/ern030.

Iglesias, R., Pérez, Y., De Torre, C., Ferreras, J. M., Antolín, P., Jiménez, P., et al. (2005). Molecular characterization and systemic induction of single-chain ribosome-inactivating proteins (RIPs) in sugar beet (Beta vulgaris) leaves. *J. Exp. Bot.* 56, 1675–1684. doi:10.1093/jxb/eri164.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223. doi:10.1126/science.1168978.

Irvin, J. (1975). Purification and partial characterization of the antiviral protein from Phytolacca americana which inhibits eukaryotic protein synthesis. *Arch Biochem Biophys* 169, 522–528.

Irvin, J. D., Kelly, T., and Robertus, J. D. (1980). Purification and properties of a second antiviral protein from Phytolacca americana which inactivates eukaryotic ribosomes. *Arch. Biochem. Biophys.* 200, 418–425. doi:10.1016/0003-9861(80)90372-0.

Ishag, H. Z., Li, C., Huang, L., Sun, M. X., Ni, B., Guo, C. X., et al. (2013). Inhibition of Japanese encephalitis virus infection in vitro and in vivo by pokeweed antiviral protein. *Virus Res.* 171, 89–96. doi:10.1016/j.virusres.2012.10.032.

Islam, M. R., Kung, S. S., Kimura, Y., and Funatsu, G. (1991). N-acetyl-D-glucosamine-asparagine structure in ribosome-inactivating proteins from the seeds of Luffa cylindrica and Phytolacca americana. *Agric. Biol. Chem.* 55, 1375–1381. doi:10.1080/00021369.1991.10870763.

Islam, W., Noman, A., Qasim, M., and Wang, L. (2018). Plant Responses to Pathogen Attack: Small RNAs in Focus. *Int. J. Mol. Sci.* 19, 515. doi:10.3390/ijms19020515.

Jagadeeswaran, G., Li, Y.-F., and Sunkar, R. (2013). Redox signaling mediates the expression of a sulfate-deprivation-inducible miR395 in Arabidopsis. *Plant J.* doi:10.1111/tpj.12364.

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J. A., et al. (2017). The genome of Chenopodium quinoa. *Nature* 542, 307–312. doi:10.1038/nature21370.

Jiang, S. Y., Ramamoorthy, R., Bhalla, R., Luan, H. F., Venkatesh, P. N., Cai, M., et al. (2008). Genome-wide survey of the RIP domain family in Oryza sativa and their expression profiles under various abiotic and biotic stresses. *Plant Mol. Biol.* 67, 603–614. doi:10.1007/s11103-008-9342-4.

Jin, J., Hewezi, T., and Baum, T. J. (2011). The Arabidopsis bHLH25 and bHLH27 transcription factors contribute to susceptibility to the cyst nematode Heterodera schachtii. *Plant J.* 65, 319–328. doi:10.1111/j.1365-313X.2010.04424.x.

Jin, J., Panicker, D., Wang, Q., Kim, M. J., Liu, J., Yin, J.-L., et al. (2014). Next generation sequencing unravels the biosynthetic ability of Spearmint (Mentha spicata) peltate glandular trichomes through comparative transcriptomics. *BMC Plant Biol.* 14, 292. doi:10.1186/s12870-014-0292-5.

Jobst, K. A., Klenov, A., Neller, K. C. M., and Hudak, K. A. (2016). Effect of depurination on cellular and viral RNA. In: *RNA Technologies: Modified Nucleic Acids in Biology and Medicine.* J. Barciszewski, V.A. Erdmann and S. Jurga eds., Springer Publishing Co., Inc., New York, NY. ISBN 978-3-319-34175-0. pp 273-297.

Johnston, J. S., Bennett, M. D., Rayburn, A. L., Galbraith, D. W., and Price, H. J. (1999). Reference standards for determination of DNA content of plant nuclei. *Am. J. Bot.* 86, 609–613. doi:10.2307/2656569.

Kaplan, B., Davydov, O., Knight, H., Galon, Y., Knight, M. R., Fluhr, R., et al. (2006). Rapid Transcriptome Changes Induced by Cytosolic Ca2+ Transients Reveal ABRE-Related Sequences as Ca2+-Responsive cis Elements in Arabidopsis. *Plant Cell* doi:10.1105/tpc.106.042713.

Karran, R. A., and Hudak, K. A. (2008). Depurination within the intergenic region of Brome mosaic virus RNA3 inhibits viral replication in vitro and in vivo. *Nucleic Acids Res.* 36, 7230–7239. doi:10.1093/nar/gkn896.

Kataoka, J., Habuka, N., Masuta, C., Miyano, M., and Koiwai, A. (1992). Isolation and analysis

of a genomic clone encoding a pokeweed antiviral protein. *Plant Mol. Biol.* 20, 879–886.

Kawade, K., and Masuda, K. (2009). Transcriptional control of two ribosome-inactivating protein genes expressed in spinach (Spinacia oleracea) embryos. *Plant Physiol. Biochem.* 47, 327–334. doi:10.1016/j.plaphy.2008.12.020.

Kawashima, C. G., Yoshimoto, N., Maruyama-Nakashita, A., Tsuchiya, Y. N., Saito, K., Takahashi, H., et al. (2009). Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant J.* 57, 313–21. doi:10.1111/j.1365-313X.2008.03690.x.

Khraiwesh, B., Zhu, J.-K., and Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta* 1819, 137–48. doi:10.1016/j.bbagrm.2011.05.001.

Kidner, C. A., and Martienssen, R. A. (2004). Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature* 428, 81–84. doi:10.1038/nature02366.

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth.3317.

Kim, H. J., Kim, Y. K., Park, J. Y., and Kim, J. (2002). Light signalling mediated by phytochrome plays an important role in cold-induced gene expression through the C-repeat/dehydration responsive element (C/DRE) in Arabidopsis thaliana. *Plant J.* doi:10.1046/j.1365-313X.2002.01249.x.

Kim, M. J., Kim, H., Shin, J. S., Chung, C. H., Ohlrogge, J. B., and Suh, M. C. (2006). Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative cis-regulatory elements in the SeFAD2 promoter and enhancers in the 5′-UTR intron. *Mol. Genet. Genomics* 276, 351–368. doi:10.1007/s00438-006-0148-2.

Kim, S. Y., Chung, H. J., and Thomas, T. L. (1997). Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the Dc3 promoter using a modified yeast one-hybrid system. *Plant J.* doi:10.1046/j.1365-313X.1997.11061237.x.

Kizis, D., and Pagès, M. (2002). Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. *Plant J.* 30, 679–689. doi:10.1046/j.1365-313X.2002.01325.x.

Klenov, A., Neller, K. C. M., Burns, L. A., Krivdova, G., and Hudak, K. A. (2015). A small RNA targets pokeweed antiviral protein transcript. *Physiol. Plant.* doi:10.1111/ppl.12393.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5, 59. doi:10.1186/1471-2105-5-59.

Korth, K. L. (2006). Medicago truncatula Mutants Demonstrate the Role of Plant Calcium Oxalate Crystals as an Effective Defense against Chewing Insects. *Plant Physiol.* 141, 188–195. doi:10.1104/pp.106.076737.

Kovalenko, T. F., and Patrushev, L. I. (2018). Pseudogenes as Functionally Significant Elements

of the Genome. *Biochem.* 83, 1332–1349. doi:10.1134/s0006297918110044.

Ku, A., Tran, D. H. T., Winge, P., Jørstad, T. S., Reese, J. C., and Troczy, J. (2011). Testing the importance of jasmonate signalling in induction of plant defences upon cabbage aphid (Brevicoryne brassicae) attack. *BMC Genomics* 12, 423. doi:10.1186/1471-2164-12-423.

Ku, Y.-S., Sintaha, M., Cheung, M.-Y., and Lam, H.-M. (2018). Plant Hormone Signaling Crosstalks between Biotic and Abiotic Stress Responses. *Int. J. Mol. Sci.* 19, 3206. doi:10.3390/ijms19103206.

Kurinov, I. V., and Uckun, F. M. (2003). High resolution X-ray structure of potent anti-HIV pokeweed antiviral protein-III. *Biochem. Pharmacol.* 65, 1709–1717. doi:10.1016/S0006-2952(03)00144-8.

Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., and Strömvik, M. V. (2018). Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.01660.

Laloi, C., Mestres-Ortega, D., Marco, Y., Meyer, Y., and Reichheld, J. (2004). The Arabidopsis Cytosolic Thioredoxin h5 Gene Induction by Oxidative Stress and Its W-Box-Mediated Response to Pathogen Elicitor. *Plant Physiol.* 134, 1006–1016. doi:10.1104/pp.103.035782.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi:10.1186/gb-2009-10-3-r25.

Laurie-Berry, N., Joardar, V., Street, I. H., and Kunkel, B. N. (2006). The Arabidopsis thaliana JASMONATE INSENSITIVE 1 gene is required for suppression of salicylic acid-dependent defenses during infection by Pseudomonas syringae. *Mol. Plant. Microbe. Interact.* 19, 789–800. doi:10.1094/MPMI-19-0789.

Laxa, M. (2017). Intron-Mediated Enhancement: A Tool for Heterologous Gene Expression in Plants? *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01977.

Lee, C. H., and Carroll, B. J. (2018). Evolution and Diversification of Small RNA Pathways in Flowering Plants. *Plant Cell Physiol.* 59, 2169–2187. doi:10.1093/pcp/pcy167.

Lee, H., Golicz, A. A., Bayer, P. E., Jiao, Y., Tang, H., Paterson, A. H., et al. (2016). The Genome of a Southern Hemisphere Seagrass Species ( *Zostera muelleri* ). *Plant Physiol.* 172, 272–283. doi:10.1104/pp.16.00868.

Lee, J. H., Wood, T. K., and Lee, J. (2015). Roles of indole as an interspecies and interkingdom signaling molecule. *Trends Microbiol.* 23, 707–718. doi:10.1016/j.tim.2015.08.001.

Lee, S. C., Lim, C. W., Lan, W., He, K., and Luan, S. (2013). ABA signaling in guard cells entails a dynamic protein-protein interaction relay from the PYL-RCAR family receptors to ion channels. *Mol. Plant* 6, 528–538. doi:10.1093/mp/sss078.

Lehmann, K., Hause, B., Altmann, D., and Köck, M. (2001). Tomato ribonuclease LX with the functional endoplasmic reticulum retention motif HDEF is expressed during programmed cell death processes, including xylem differentiation, germination, and senescence. *Plant Physiol.* 127, 436–449. doi:10.1104/pp.010362.

Lei, M., La, H., Lu, K., Wang, P., Miki, D., Ren, Z., et al. (2013). Arabidopsis EDM2 promotes IBM1 distal polyadenylation and regulates genome DNA methylation patterns. *Proc. Natl. Acad. Sci.* 111, 527–532. doi:10.1073/pnas.1320106110.

Lers, A., Sonego, L., Green, P. J., and Burd, S. (2006). Suppression of LX ribonuclease in tomato results in a delay of leaf senescence and abscission. *Plant Physiol.* 142, 710–721. doi:10.1104/pp.106.080135.

Levine, E., McHale, P., and Levine, H. (2007). Small regulatory RNAs may sharpen spatial expression patterns. *PLoS Comput. Biol.* 3, 2356–2365. doi:10.1371/journal.pcbi.0030233.

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. doi:10.1186/1471-2105-12-323.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. *Curr. Biol.* 15, 1501–1507. doi:10.1016/j.cub.2005.07.029.

Li, S., Castillo-González, C., Yu, B., and Zhang, X. (2017). The functions of plant small RNAs in development and in stress responses. *Plant J.* 90, 654–670. doi:10.1111/tpj.13444.

Li, X., Lu, J., Liu, S., Liu, X., Lin, Y., and Li, L. (2014). Identification of rapidly induced genes in the response of peanut (Arachis hypogaea) to water deficit and abscisic acid. *BMC Biotechnol.* 14, 58. doi:10.1186/1472-6750-14-58.

Ling, M. H. T., Ban, Y., Wen, H., Wang, S. M., and Ge, S. X. (2013). Conserved expression of natural antisense transcripts in mammals. *BMC Genomics* 14, 243. doi:10.1186/1471-2164-14-243.

Liu, J.-J., Sturrock, R. N., and Benton, R. (2013). Transcriptome analysis of Pinus monticola primary needles by RNA-seq provides novel insight into host resistance to Cronartium ribicola. *BMC Genomics* 14, 884. doi:10.1186/1471-2164-14-884.

Liu, Q., Luo, L., and Zheng, L. (2018). Lignins: Biosynthesis and biological functions in plants. *Int. J. Mol. Sci.* 19. doi:10.3390/ijms19020335.

Liu, X., Hao, L., Li, D., Zhu, L., and Hu, S. (2015). Long Non-coding RNAs and Their Biological Roles in Plants. *Genomics. Proteomics Bioinformatics*. doi:10.1016/j.gpb.2015.02.003.

Liu, X., Peng, K., Wang, A., Lian, C., and Shen, Z. (2010). Cadmium accumulation and distribution in populations of Phytolacca americana L. and the role of transpiration. *Chemosphere* 78, 1136–1141. doi:10.1016/j.chemosphere.2009.12.030.

Liu, Y., Ahn, J.-E., Datta, S., Salzman, R. A., Moon, J., Huyghues-Despointes, B., et al. (2005). Arabidopsis vegetative storage protein is an anti-insect acid phosphatase. *Plant Physiol.* 139, 1545–1556. doi:10.1104/pp.105.066837.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*. doi:10.1006/meth.2001.1262.

Lodge, J. K., Kaniewski, W. K., and Tumer, N. E. (1993). Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. *Proc. Natl. Acad. Sci. U. S. A.* 90, 7089–7093. doi:10.1073/pnas.90.15.7089.

Lopez-Gomollon, S., Mohorianu, I., Szittya, G., Moulton, V., and Dalmay, T. (2012). Diverse correlation patterns between microRNAs and their targets during tomato fruit development indicates different modes of microRNA actions. *Planta* 236, 1875–1887. doi:10.1007/s00425-012-1734-7.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26. doi:10.1186/1748-7188-6-26.

Lorenzo, O., Chico, J. M., Sánchez-Serrano, J. J., and Solano, R. (2004). JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis. *Plant Cell* 16, 1938–1950. doi:10.1105/tpc.022319.

Love, R. R., Weisenfeld, N. I., Jaffe, D. B., Besansky, N. J., and Neafsey, D. E. (2016). Evaluation of DISCOVAR de novo using a mosquito sample for cost-effective short-read genome assembly. *BMC Genomics* 17, 187. doi:10.1186/s12864-016-2531-7.

Lu, Z., Feng, X., Song, L., Han, Y., Kim, A., Herzberg, O., et al. (2005). Diversity of function in the isocitrate lyase enzyme superfamily: The Dianthus caryophyllus petal death protein cleaves α-keto and α-hydroxycarboxylic acids. *Biochemistry* 44, 16365–16376. doi:10.1021/bi051776l.

Luo, H., Song, F., Goodman, R. M., and Zheng, Z. (2005). Up-regulation of OsBIHD1, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses. *Plant Biol.* 7, 459–468. doi:10.1055/s-2005-865851.

Ma, J., Hanssen, M., Lundgren, K., Hernández, L., Delatte, T., Ehlert, A., et al. (2011). The sucrose-regulated Arabidopsis transcription factor bZIP11 reprograms metabolism and regulates trehalose metabolism. *New Phytol.* 191, 733–745. doi:10.1111/j.1469-8137.2011.03735.x.

Ma, L., and Li, G. (2018). FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) Family Proteins in Arabidopsis Growth and Development. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.00692.

Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., et al. (2009). Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* 324, 1064–1068. doi:10.1126/science.1172408.

Maldonado, A. M., Doerner, P., Dixonk, R. A., Lamb, C. J., and Cameron, R. K. (2002). A putative lipid transfer protein involved in systemic resistance signalling in Arabidopsis. *Nature* 419, 399–403. doi:10.1038/nature00962.

Manners, J. M., Penninckx, I. A. M. A., Vermaere, K., Kazan, K., Brown, R. L., Morgan, A., et al. (1998). The promoter of the plant defensin gene PDF1.2 from Arabidopsis is systemically activated by fungal pathogens and responds to methyl jasmonate but not to

salicylic acid. *Plant Mol. Biol.* 38, 1071–1080. doi:10.1023/A:1006070413843.

Mansouri, S., Choudhary, G., Sarzala, P. M., Ratner, L., and Hudak, K. A. (2009). Suppression of human T-cell leukemia virus I gene expression by pokeweed antiviral protein. *J. Biol. Chem.* 284, 31453–31462. doi:10.1074/jbc.M109.046235.

Marcotte, W. R., Russell, S. H., and Quatrano, R. S. (1989). Abscisic Acid-Responsive Sequences from the Em Gene of Wheat. *Plant Cell* 1, 969. doi:10.2307/3868997.

Maruyama-Nakashita, A., Nakamura, Y., Watanabe-Takahashi, A., Inoue, E., Yamaya, T., and Takahashi, H. (2005). Identification of a novel cis-acting element conferring sulfur deficiency response in Arabidopsis roots. *Plant J.* 42, 305–314. doi:10.1111/j.1365-313X.2005.02363.x.

Marzec, M. (2016). Strigolactones as Part of the Plant Defence System. *Trends Plant Sci.* 21, 900–903. doi:10.1016/j.tplants.2016.08.010.

Matthewman, C. A., Kawashima, C. G., Húska, D., Csorba, T., Dalmay, T., and Kopriva, S. (2012). MiR395 is a general component of the sulfate assimilation regulatory network in Arabidopsis. *FEBS Lett.* 586, 3242–3248. doi:10.1016/j.febslet.2012.06.044.

Mena, M., Cejudo, F., Isabel-Lamoneda, I., and Carbonero, P. (2002). A Role for the DOF Transcription Factor BPBF in the Regulation of Gibberellin-Responsive Genes in Barley Aleurone. *Plant Physiol.* doi:10.1104/pp.005561.

Mohanty, B., Krishnan, S. P. T., Swarup, S., and Bajic, V. B. (2005). Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. *Ann. Bot.* doi:10.1093/aob/mci219.

Chávez Montes, R. A., De Fátima Rosas-Cárdenas, F., De Paoli, E., Accerbi, M., Rymarquis, L. A., Mahalingam, G., et al. (2014). Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat. Commun.* 5, 3722. doi:10.1038/ncomms4722.

Morello, L., Bardini, M., Cricrì, M., Sala, F., and Breviario, D. (2006). Functional analysis of DNA sequences controlling the expression of the rice OsCDPK2 gene. *Planta* 223, 479–491. doi:10.1007/s00425-005-0105-z.

Morello, L., Bardini, M., Sala, F., and Breviario, D. (2002). A long leader intron of the *Ostub16* rice beta-tubulin gene is required for high-level gene expression and can autonomously promote transcription both *in vivo* and *in vitro*. *Plant J.* 29, 33–44. doi:10.1046/j.0960-7412.2001.01192.x.

Moreton, J., Izquierdo, A., and Emes, R. D. (2016). Assembly, assessment, and availability of De novo generated eukaryotic transcriptomes. *Front. Genet.* 6. doi:10.3389/fgene.2015.00361.

Mroue, S., Simeunovic, A., and Robert, H. S. (2018). Auxin production as an integrator of environmental cues for developmental growth regulation. *J. Exp. Bot.* 69, 201–212. doi:10.1093/jxb/erx259.

Muniategui, A., Pey, J., Planes, F. J., and Rubio, A. (2013). Joint analysis of miRNA and mRNA expression data. *Brief. Bioinform.* 14, 263–78. doi:10.1093/bib/bbs028.

Mur, L. A. J. (2005). The Outcomes of Concentration-Specific Interactions between Salicylate and Jasmonate Signaling Include Synergy, Antagonism, and Oxidative Stress Leading to Cell Death. *Plant Physiol.* 140, 249–262. doi:10.1104/pp.105.072348.

Mutasa-Göttgens, E. S., Joshi, A., Holmes, H. F., Hedden, P., and Göttgens, B. (2012). A new RNASeq-based reference transcriptome for sugar beet and its application in transcriptome-scale analysis of vernalization and gibberellin responses. *BMC Genomics* 13, 99. doi:10.1186/1471-2164-13-99.

Neller, K. C. M., Klenov, A., Guzman, J. C., and Hudak, K. A. (2018). Integration of the Pokeweed miRNA and mRNA Transcriptomes Reveals Targeting of Jasmonic Acid-Responsive Genes. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.00589.

Neller, K. C. M., Klenov, A., and Hudak, K. A. (2016). The Pokeweed Leaf mRNA Transcriptome and Its Regulation by Jasmonic Acid. *Front. Plant Sci.* 7, 1–13. doi:10.3389/fpls.2016.00283.

Neller, K. C. M., Klenov, A., and Hudak, K. A. (2019). Prediction and Characterization of miRNA/Target Pairs in Non-Model Plants Using RNA-seq. *Curr. Protoc. Plant Biol.*, e20090. doi:10.1002/cppb.20090.

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., et al. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. *Plant Cell* 18, 2929–2945. doi:10.1105/tpc.106.045617.

Nishiuchi, T., Shinshi, H., and Suzuki, K. (2004). Rapid and transient activation of transcription of the ERF3 gene by wounding in tobacco leaves: Possible involvement of NtWRKYs and autorepression. *J. Biol. Chem.* doi:10.1074/jbc.M409674200.

Oeder, S., Mages, J., Flicek, P., and Lang, R. (2007). Uncovering information on expression of natural antisense transcripts in Affymetrix MOE430 datasets. *BMC Genomics* 8, 200. doi:10.1186/1471-2164-8-200.

Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y., and Yamaguchi, S. (2003). Gibberellin Biosynthesis and Response during Arabidopsis Seed Germination. *Plant Cell*. doi:10.1105/tpc.011650.

Okuda, K., Chateigner-Boutin, A.-L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., et al. (2009). Pentatricopeptide Repeat Proteins with the DYW Motif Have Distinct Molecular Functions in RNA Editing and RNA Cleavage in Arabidopsis Chloroplasts. *Plant Cell* 21, 146–156. doi:10.1105/tpc.108.064667.

Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc. Natl. Acad. Sci.* 104, 8178–8183. doi:10.1073/pnas.0700865104.

Pandey, S. P., Shahi, P., Gase, K., and Baldwin, I. T. (2008). Herbivory-induced changes in the small-RNA transcriptome and phytohormone signaling in Nicotiana attenuata. *Proc. Natl. Acad. Sci.* 105, 4559–4564. doi:10.1073/pnas.0711363105.

Parente, A., Conforto, B., Di Maro, A., Chambery, A., De Luca, P., Bolognesi, A., et al. (2008). Type 1 ribosome-inactivating proteins from Phytolacca dioica L. leaves: differential

seasonal and age expression, and cellular localization. *Planta* 228, 963–975. doi:10.1007/s00425-008-0796-z.

Parikh, B. A., Coetzer, C., and Tumer, N. E. (2002). Pokeweed antiviral protein regulates the stability of its own mRNA by a mechanism that requires depurination but can be separated from depurination of the alpha-sarcin/ricin loop of rRNA. *J. Biol. Chem.* 277, 41428–37. doi:10.1074/jbc.M205463200.

Park, H., Kim, M., Kang, Y., Jeon, J., Yoo, J., Kim, M., et al. (2004). Pathogen- and NaCl-Induced Expression of the SCaM-4 Promoter Is Mediated in Part by a GT-1 Box That Interacts with a GT-1-Like Transcription Factor. *Plant Physiol.* doi:10.1104/pp.104.041442.

Park, S.-Y., Fung, P., Nishimura, N., Jensen, D. R., Fujii, H., Zhao, Y., et al. (2009). Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* 324, 1068–1071. doi:10.1126/science.1173041.

Park, S., Lawrence, C. B., Linden, J. C., and Vivanco, J. M. (2002). Isolation and Characterization of a Novel Ribosome- Inactivating Protein from Root Cultures of Pokeweed and Its Mechanism of Secretion from Roots. *Plant Physiol.* 130, 164–178. doi:10.1104/pp.000794.of.

Parthasarathy, A., Cross, P. J., Dobson, R. C. J., Adams, L. E., Savka, M. A., and Hudson, A. O. (2018). A Three-Ring circus: Metabolism of the three proteogenic aromatic amino acids and their role in the health of plants and animals. *Front. Mol. Biosci.* 5. doi:10.3389/fmolb.2018.00029.

Pelechano, V., and Steinmetz, L. M. (2013). Gene regulation by antisense transcription. *Nat. Rev. Genet.* 14, 880–93. doi:10.1038/nrg3594.

Peng, K., Luo, C., You, W., Lian, C., Li, X., and Shen, Z. (2008). Manganese uptake and interactions with cadmium in the hyperaccumulator-Phytolacca Americana L. *J. Hazard. Mater.* 154, 674–681. doi:10.1016/j.jhazmat.2007.10.080.

Peng, Z., He, S., Gong, W., Sun, J., Pan, Z., Xu, F., et al. (2014). Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. *BMC Genomics* 15, 1–28. doi:10.1186/1471-2164-15-760.

Penninckx, I. A., Thomma, B. P., Buchala, A., Métraux, J. P., and Broekaert, W. F. (1998). Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in Arabidopsis. *Plant Cell* 10, 2103–2113. doi:10.1105/tpc.10.12.2103.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122.

Pierce, E. J., and Rey, M. E. C. (2013). Assessing Global Transcriptome Changes in Response to South African Cassava Mosaic Virus [ZA-99] Infection in Susceptible Arabidopsis thaliana. *PLoS One* 8. doi:10.1371/journal.pone.0067534.

Pieterse, C. M., van Wees, S. C., van Pelt, J. A., Knoester, M., Laan, R., Gerrits, H., et al. (1998). A novel signaling pathway controlling induced systemic resistance in Arabidopsis. *Plant*

*Cell* 10, 1571–1580.

Polashock, J., Zelzion, E., Fajardo, D., Zalapa, J., Georgi, L., Bhattacharya, D., et al. (2014). The American cranberry: First insights into the whole genome of a species adapted to bog habitat. *BMC Plant Biol.* 14. doi:10.1186/1471-2229-14-165.

Poyet, J. (1997). Presence of an Intron in a Gene of PAP II, the Ribosome-inactivating Protein from Summer Leaves of Phytolacca americana. *Ann. Bot.* 80, 685–688. doi:10.1006/anbo.1997.0478.

Pundir, S., Martin, M. J., and O'Donovan, C. (2017). "UniProt protein knowledgebase," in *Methods in Molecular Biology*, 41–55. doi:10.1007/978-1-4939-6783-4_2.

Qi, X. T., Zhang, Y. X., and Chai, T. Y. (2007). The bean PvSR2 gene produces two transcripts by alternative promoter usage. *Biochem. Biophys. Res. Commun.* 356, 273–278. doi:10.1016/j.bbrc.2007.02.124.

Qin, X., Shao, C., Hou, P., Gao, J., Lei, N., and Jiang, L. (2010). Different Functions and Expression Profiles of Curcin and Curcin-L in Jatropha curcas L. *Z Naturforsch C*. 65, 355-62.

Qin, X., Zheng, X., Shao, C., Gao, J., Jiang, L., Zhu, X., et al. (2009). Stress-induced curcin-L promoter in leaves of Jatropha curcas L. and characterization in transgenic tobacco. *Planta* 230, 387–395. doi:10.1007/s00425-009-0956-9.

Qiu, D., Pan, X., Wilson, I. W., Li, F., Liu, M., Teng, W., et al. (2009). High throughput sequencing technology reveals that the taxoid elicitor methyl jasmonate regulates microRNA expression in Chinese yew (Taxus chinensis). *Gene* 436, 37–44. doi:10.1016/j.gene.2009.01.006.

R Development Core Team, R. (2011). R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* 1, 409. doi:10.1007/978-3-540-74686-7.

R Foundation for Statistical Computing (2016). R: A Language and Environment for Statistical Computing. in *R Foundation for Statistical Computing* doi:10.1007/978-3-540-74686-7.

Rajamohan, F., Venkatachalam, T. K., Irvin, J. D., and Uckun, F. M. (1999). Pokeweed antiviral protein isoforms PAP-I, PAP-II, and PAP-III depurinate RNA of human immunodeficiency virus (HIV)-1. *Biochem. Biophys. Res. Commun.* 260, 453–458. doi:10.1006/bbrc.1999.0922.

Ranjan, A., Ichihashi, Y., Farhi, M., Zumstein, K., Townsley, B., David-Schwartz, R., et al. (2014). De novo assembly and characterization of the transcriptome of the parasitic weed dodder identifies genes associated with plant parasitism. *Plant Physiol.* 166, 1186–99. doi:10.1104/pp.113.234864.

Ranocha, P., Denancé, N., Vanholme, R., Freydier, A., Martinez, Y., Hoffmann, L., et al. (2010). Walls are thin 1 (WAT1), an Arabidopsis homolog of Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J.* 63, 469–483. doi:10.1111/j.1365-313X.2010.04256.x.

Ready, M. P., Brown, D. T., and Robertus, J. D. (1986). Extracellular localization of pokeweed

antiviral protein. *Proc. Natl. Acad. Sci. U. S. A.* 83, 5053–5056. doi:10.1073/pnas.83.14.5053.

Reinbothe, S., Reinbothe, C., Lehmann, J., Becker, W., Apel, K., and Parthier, B. (1994). JIP60, a methyl jasmonate-induced ribosome-inactivating protein involved in plant stress reactions. *Proc Natl Acad Sci U S A* 91, 7012–7016.

Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., et al. (2015). The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers. *New Phytol.* 206, 19–26. doi:10.1111/nph.13191.

Rieping, M., and Schöffl, F. (1992). Synergistic effect of upstream sequences, CCAAT box elements, and HSE sequences for enhanced expression of chimaeric heat shock genes in transgenic tobacco. *Mol. Gen. Genet.* 231, 226–232. doi:10.1007/BF00279795.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–40. doi:10.1093/bioinformatics/btp616.

Rosikiewicz, W., and Makałowska, I. (2016). Biological functions of natural antisense transcripts. *Acta Biochim. Pol.* 63, 665–673. doi:10.18388/abp.2016_1350.

Rouster, J., Leah, R., Mundy, J., and Cameron-Mills, V. (1997). Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. *Plant J.* 11, 513–523. doi:10.1046/j.1365-313X.1997.11030513.x.

Salminen, T. A., Blomqvist, K., and Edqvist, J. (2016). Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta* 244, 971–997. doi:10.1007/s00425-016-2585-4.

Satoh, R., Fujita, Y., Nakashima, K., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). A novel subgroup of bZIP proteins functions as transcriptional activators in hypoosmolarity-responsive expression of the ProDH gene in Arabidopsis. *Plant Cell Physiol.* 45, 309–317. doi:10.1093/pcp/pch036.

Schaller, A., and Stintzi, A. (2009). Enzymes in jasmonate biosynthesis - Structure, function, regulation. *Phytochemistry* 70, 1532–1538. doi:10.1016/j.phytochem.2009.07.032.

Schluttenhofer, C., Pattanaik, S., Patra, B., and Yuan, L. (2014). Analyses of Catharanthus roseus and Arabidopsis thaliana WRKY transcription factors reveal involvement in jasmonate signaling. *BMC Genomics* 15. doi:10.1186/1471-2164-15-502.

Schommer, C., Palatnik, J. F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E. E., et al. (2008). Control of jasmonate biosynthesis and senescence by miR319 targets. *PLoS Biol.* 6, e230. doi:10.1371/journal.pbio.0060230.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–504. doi:10.1101/gr.1239303.

Shaul, O. (2017). How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* 91, 145–155. doi:10.1016/j.biocel.2017.06.016.

Shen, E. M., Singh, S. K., Ghosh, J. S., Patra, B., Paul, P., Yuan, L., et al. (2017). The miRNAome of Catharanthus roseus: identification, expression analysis, and potential roles of microRNAs in regulation of terpenoid indole alkaloid biosynthesis. *Sci. Rep.* 7, 43027. doi:10.1038/srep43027.

Shi, H. (2003). The Arabidopsis SOS5 Locus Encodes a Putative Cell Surface Adhesion Protein and Is Required for Normal Cell Expansion. *Plant Cell* 15, 19–32. doi:10.1105/tpc.007872.

Shuai, P., Liang, D., Zhang, Z., Yin, W., and Xia, X. (2013). Identification of drought-responsive and novel Populus trichocarpa microRNAs by high- throughput sequencing and their targets using degradome analysis. *BMC Genomics.* 14, 233.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: user guide. *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351.

Simpson, S. D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. *Plant J.* doi:10.1046/j.1365-313X.2003.01624.x.

Singh, I., and Shah, K. (2014). Exogenous application of methyl jasmonate lowers the effect of cadmium-induced oxidative injury in rice seedlings. *Phytochemistry* 108, 57–66. doi:10.1016/j.phytochem.2014.09.007.

Smith, C. D., Edgar, R. C., Yandell, M. D., Smith, D. R., Celniker, S. E., Myers, E. W., et al. (2007). Improved repeat identification and masking in Dipterans. *Gene* 389, 1–9. doi:10.1016/j.gene.2006.09.011.

Söderman, E., Mattsson, J., and Engström, P. (1996). The Arabidopsis homeobox gene ATHB-7 is induced by water deficit and by abscisic acid. *Plant J.* 10, 375–381. doi:10.1046/j.1365-313X.1996.10020375.x.

Song, Q. A., Catlin, N. S., Brad Barbazuk, W., and Li, S. (2019). Computational analysis of alternative splicing in plant genomes. *Gene* 685, 186–195. doi:10.1016/j.gene.2018.10.026.

Song, S. K., Choi, Y., Moon, Y. H., Kim, S. G., Choi, Y. D., and Lee, J. S. (2000). Systemic induction of a Phytolacca insularis antiviral protein gene by mechanical wounding, jasmonic acid, and abscisic acid. *Plant Mol. Biol.* 43, 439–50.

Spoel, S. H., Koornneef, A., Claessens, S. M. C., Korzelius, J. P., Van Pelt, J. A., Mueller, M. J., et al. (2003). NPR1 modulates cross-talk between salicylate- and jasmonate-dependent defense pathways through a novel function in the cytosol. *Plant Cell* 15, 760–770. doi:10.1105/tpc.009159.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32. doi:10.1093/nar/gkh379.

Stintzi, A., and Browse, J. (2000). The Arabidopsis male-sterile mutant, opr3, lacks the 12-oxophytodienoic acid reductase required for jasmonate synthesis. *Proc. Natl. Acad. Sci.* 97, 10625–10630. doi:10.1073/pnas.190264497.

Stirpe, F. (2013). Ribosome-inactivating proteins: From toxins to useful proteins. *Toxicon* 67,

12–16. doi:10.1016/j.toxicon.2013.02.005.

Stocks, M. B., Moxon, S., Mapleson, D., Woolfenden, H. C., Mohorianu, I., Folkes, L., et al. (2012). The UEA sRNA workbench: A suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics* 28, 2059–2061. doi:10.1093/bioinformatics/bts311.

Sun, Y. C., Pan, L. L., Ying, F. Z., Li, P., Wang, X. W., and Liu, S. S. (2017). Jasmonic acid-related resistance in tomato mediates interactions between whitefly and whitefly-transmitted virus. *Sci. Rep.* 7. doi:10.1038/s41598-017-00692-w.

Sutoh, K., and Yamauchi, D. (2003). Two cis-acting elements necessary and sufficient for gibberellin-upregulated proteinase expression in rice seeds. *Plant J.* doi:10.1046/j.1365-313X.2003.01753.x.

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007). UniRef: Comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23, 1282–1288. doi:10.1093/bioinformatics/btm098.

Tada, Y., Spoel, S. H., Pajerowska-Mukhtar, K., Mou, Z., Song, J., Wang, C., et al. (2008). Plant immunity requires conformational changes [corrected] of NPR1 via S-nitrosylation and thioredoxins. *Science* 321, 952–956. doi:10.1126/science.1156970.

Tarkowská, D., and Strnad, M. (2018). Isoprenoid-derived plant signaling molecules: biosynthesis and biological importance. *Planta* 247, 1051–1066. doi:10.1007/s00425-018-2878-x.

Tartarini, A., Pittaluga, E., Marcozzi, G., Testone, G., Rodrigues-Pousada, R. A., Giannino, D., et al. (2010). Differential expression of saporin genes upon wounding, ABA treatment and leaf development. *Physiol. Plant.* 140, 141–152. doi:10.1111/j.1399-3054.2010.01388.x.

Tehseen, M., Cairns, N., Sherson, S., and Cobbett, C. S. (2010). Metallochaperone-like genes in Arabidopsis thaliana. *Metallomics* 2, 556–564. doi:10.1039/c003484c.

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., et al. (2007). JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. *Nature* 448, 661–665. doi:10.1038/nature05960.

Tourlakis, M. E., Karran, R. A., Desouza, L., Siu, K. W. M., and Hudak, K. A. (2010). Homodimerization of pokeweed antiviral protein as a mechanism to limit depurination of pokeweed ribosomes. *Mol. Plant Pathol.* 11, 757–767. doi:10.1111/j.1364-3703.2010.00640.x.

Tsang, J., Zhu, J., and van Oudenaarden, A. (2007). MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. *Mol. Cell* 26, 753–767. doi:10.1016/j.molcel.2007.05.018.

Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F., Yamaguchi-Shinozaki, K., et al. (2009). Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 106, 17588–17593. doi:10.1073/pnas.0907095106.

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2000). Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11632–11637. doi:10.1073/pnas.190309197.

Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Taniai, N., Miyagi, N., et al. (2017). Draft genome sequence of bitter gourd (Momordica charantia), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res.* 24, 51–58. doi:10.1093/dnares/dsw047.

Van Hoeck, A., Horemans, N., Monsieurs, P., Cao, H. X., Vandenhove, H., and Blust, R. (2015). The first draft genome of the aquatic model plant Lemna minor opens the route for future stress physiology research and biotechnological applications. *Biotechnol. Biofuels* 8. doi:10.1186/s13068-015-0381-1.

Veeckman, E., Ruttink, T., and Vandepoele, K. (2016). Are We There Yet? Reliably Estimating the Completeness of Plant Genome Sequences. *Plant Cell* 28, 1759–1768. doi:10.1105/tpc.16.00349.

Vepachedu, R., Bais, H. P., and Vivanco, J. M. (2003). Molecular characterization and post-transcriptional regulation of ME1, a type-I ribosome-inactivating protein from Mirabilis expansa. *Planta* 217, 498–506. doi:10.1007/s00425-003-1014-7.

Verma, V., Ravindran, P., and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 16. doi:10.1186/s12870-016-0771-y.

Vivanco, J. M., and Tumer, N. E. (2003). Translation Inhibition of Capped and Uncapped Viral RNAs Mediated by Ribosome-Inactivating Proteins. *Phytopathology* 93, 588–595. doi:10.1094/phyto.2003.93.5.588.

Wan, L., Wang, F., Guo, X., Lu, S., Qiu, Z., Zhao, Y., et al. (2012). Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing. *BMC Plant Biol.* 12, 146.

Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J., et al. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome Res.* 24, 444–453. doi:10.1101/gr.165555.113.

Wang, L., Li, Q., Zhang, A., Zhou, W., Jiang, R., Yang, Z., et al. (2017). The Phytol Phosphorylation Pathway Is Essential for the Biosynthesis of Phylloquinone, which Is Required for Photosystem I Stability in Arabidopsis. *Mol. Plant* 10, 183–196. doi:10.1016/j.molp.2016.12.006.

Wang, M., and Hudak, K. A. (2006). A novel interaction of pokeweed antiviral protein with translation initiation factors 4G and iso4G: A potential indirect mechanism to access viral RNAs. *Nucleic Acids Res.* 34, 1174–1181. doi:10.1093/nar/gkj520.

Wang, P., Zoubenko, O., and Tumer, N. E. (1998). Reduced toxicity and broad spectrum resistance to viral and fungal infection in transgenic plants expressing pokeweed antiviral protein II. *Plant Mol. Biol.* 38, 957–64.

Wang, W., Tang, W., Ma, T., Niu, D., Jin, J. B., Wang, H., et al. (2016). A pair of light signaling factors FHY3 and FAR1 regulates plant immunity by modulating chlorophyll biosynthesis.

*J. Integr. Plant Biol.* 58, 91–103. doi:10.1111/jipb.12369.

Wasternack, C., and Feussner, I. (2017). The Oxylipin Pathways : Biochemistry and Function. *Annu. Rev. Plant Biol.* 69, 363-386.

Wasternack, C., and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Ann. Bot.* 111, 1021–58. doi:10.1093/aob/mct067.

Wasternack, C., and Song, S. (2017). Jasmonates: Biosynthesis, metabolism, and signaling by proteins activating and repressing transcription. *J. Exp. Bot.* 68, 1303–1321. doi:10.1093/jxb/erw443.

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi:10.1093/molbev/msx319.

Wen, M., Xie, M., He, L., Wang, Y., Shi, S., and Tang, T. (2016). Expression Variations of miRNAs and mRNAs in Rice (Oryza sativa). *Genome Biol. Evol.* 8, 3529–3544. doi:10.1093/gbe/evw252.

Wetzel, S., Demmers, C., and Greenwood, J. S. (1989). Seasonally fluctuating bark proteins are a potential form of nitrogen storage in three temperate hardwoods. *Planta* 178, 275–281. doi:10.1007/BF00391854.

Widhalm, J. R., Ducluzeau, A. L., Buller, N. E., Elowsky, C. G., Olsen, L. J., and Basset, G. J. C. (2012). Phylloquinone (vitamin K1) biosynthesis in plants: Two peroxisomal thioesterases of lactobacillales origin hydrolyze 1,4-dihydroxy-2-naphthoyl-coa. *Plant J.* 71, 205–215. doi:10.1111/j.1365-313X.2012.04972.x.

Wildermuth, M. C., Dewdney, J., Wu, G., and Ausubel, F. M. (2001). Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* 414, 562–565. doi:10.1038/35107108.

Wise, A. A., Liu, Z., and Binns, A. N. (2006). "Three Methods for the Introduction of Foreign DNA into Agrobacterium," in *Agrobacterium Protocols*, ed. K. Wang (Totowa, NJ: Humana Press), 47–48. doi:10.1385/1-59745-130-4:43.

Woo, W. S., and Kang, S. S. (1976). Phytolaccoside B: Triterpene glucoside from Phytolacca americana. *Phytochemistry* 15, 1315–1317. doi:10.1016/0031-9422(76)85102-3.

Wu, C. I., Shen, Y., and Tang, T. (2009). Evolution under canalization and the dual roles of microRNAs - A hypothesis. *Genome Res.* 19, 734–743. doi:10.1101/gr.084640.108.

Wu, P., Zhou, C., Cheng, S., Wu, Z., Lu, W., Han, J., et al. (2015). Integrated genome sequence and linkage map of physic nut (Jatropha curcas L.), a biodiesel plant. *Plant J.* 81, 810–821. doi:10.1111/tpj.12761.

Wu, Y., Zhang, D., Chu, J. Y., Boyle, P., Wang, Y., Brindle, I. D., et al. (2012). The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell Rep.* 1, 639–47. doi:10.1016/j.celrep.2012.05.008.

Wytynck, P., Rougé, P., and Van Damme, E. J. M. (2017). Genome-wide screening of Oryza sativa ssp. japonica and indica reveals a complex family of proteins with ribosome-inactivating protein domains. *Phytochemistry* 143, 87–97. doi:10.1016/j.phytochem.2017.07.009.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, 316–322. doi:10.1093/nar/gkr483.

Xie, Z., Kasschau, K. D., and Carrington, J. C. (2003). Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. *Curr. Biol.* 13, 784–789. doi:10.1016/S0960-9822(03)00281-1.

Xu, C., Jiao, C., Sun, H., Cai, X., Wang, X., Ge, C., et al. (2017). Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. *Nat. Commun.* 8. doi:10.1038/ncomms15275.

Xu, S., Jiang, Y., Wang, N., Xia, B., Jiang, Y., Li, X., et al. (2016). Identification and differential regulation of microRNAs in response to methyl jasmonate treatment in Lycoris aurea by deep sequencing. *BMC Genomics* 17, 789. doi:10.1186/s12864-016-2645-y.

Yamamoto, S., Nakano, T., Suzuki, K., and Shinshi, H. (2004). Elicitor-induced activation of transcription via W box-related cis-acting elements from a basic chitinase gene by WRKY transcription factors in tobacco. *Biochim. Biophys. Acta - Gene Struct. Expr.* doi:10.1016/j.bbaexp.2004.07.005.

Yan, H., Yoo, M., Koh, J., Liu, L., Chen, Y., and Acikgoz, D. (2014). Molecular Reprogramming of Arabidopsis in Response to Perturbation of Jasmonate Signaling. *J. Proteome Res.*13, 5751-66.

Yan, Z., Chen, J., and Li, X. (2013). Methyl jasmonate as modulator of Cd toxicity in Capsicum frutescens var. fasciculatum seedlings. *Ecotoxicol. Environ. Saf.* 98, 203–209. doi:10.1016/j.ecoenv.2013.08.019.

Yang, X., Feng, Y., He, Z., and Stoffella, P. J. (2005). Molecular mechanisms of heavy metal hyperaccumulation and phytoremediation. *J. Trace Elem. Med. Biol.*, 339–353. doi:10.1016/j.jtemb.2005.02.007.

Yang, X., and Li, L. (2011). miRDeep-P: A computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics* 27, 2614–2615. doi:10.1093/bioinformatics/btr430.

Yasuda, M., Ishikawa, A., Jikumaru, Y., Seki, M., Umezawa, T., Asami, T., et al. (2008). Antagonistic interaction between systemic acquired resistance and the abscisic acid-mediated abiotic stress response in Arabidopsis. *Plant Cell* 20, 1678–1692. doi:10.1105/tpc.107.054296.

Yoshinari, S., Koresawa, S., Yokota, S., Sawamoto, H., Tamura, M., and Endo, Y. (2009). Gypsophilin, a New Type 1 Ribosome-inactivating Protein from Gypsophila elegans : Purification, Enzymatic Characterization, and Subcellular Localization . *Biosci. Biotechnol. Biochem.* 61, 324–331. doi:10.1271/bbb.61.324.

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis

for RNA-seq: accounting for selection bias. *Genome Biol.* 11. doi:10.1186/gb-2010-11-2-r14.

Yu, D., Chen, C., Chen, Z., Jiang, C.-J., Ono, K., Toki, S., et al. (2001). Evidence for an Important Role of WRKY DNA Binding Proteins in the Regulation of NPR1 Gene Expression. *Plant Cell* 13, 1527–1540. doi:10.1105/TPC.010115.

Yu, Y., Jia, T., and Chen, X. (2017). The "how" and "where" of plant microRNAs. *New Phytol.* 216, 1002–1017. doi:10.1111/nph.14834.

Zander, M., La Camera, S., Lamotte, O., Métraux, J. P., and Gatz, C. (2010). Arabidopsis thaliana class-II TGA transcription factors are essential activators of jasmonic acid/ethylene-induced defense responses. *Plant J.* 61, 200–210. doi:10.1111/j.1365-313X.2009.04044.x.

Zdobnov, E. M., and Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848. doi:10.1093/bioinformatics/17.9.847.

Zeng, M., Zheng, M., Lu, D., Wang, J., Jiang, W., and Sha, O. (2015). Anti-tumor activities and apoptotic mechanism of ribosome-inactivating proteins. *Chin. J. Cancer* 34. doi:10.1186/s40880-015-0030-x.

Zhabokritsky, A., Mansouri, S., and Hudak, K. A. (2014). Pokeweed antiviral protein alters splicing of HIV-1 RNAs, resulting in reduced virus production. *RNA* 20, 1238–1247. doi:10.1261/rna.043141.113.

Zhang, B., Jin, Z., and Xie, D. (2012). Global Analysis of Non-coding Small RNAs in Arabidopsis in Response to Jasmonate Treatment by Deep Sequencing Technology. *J. Integr. Plant Biol.* 54, 73–86. doi:10.1111/j.1744-7909.2012.01098.x.

Zhang, G., Yi, H., and Gong, J. (2014). The Arabidopsis Ethylene / Jasmonic Acid-NRT Signaling Module Coordinates Nitrate Reallocation and the Trade-Off between Growth and Environmental Adaptation. *Plant Cell* 26, 3984–3998. doi:10.1105/tpc.114.129296.

Zhang, G., Zhao, F., Chen, L., Pan, Y., Sun, L., Bao, N., et al. (2019). Jasmonate-mediated wound signalling promotes plant regeneration. *Nat. Plants* 5, 491–497. doi:10.1038/s41477-019-0408-x.

Zhang, L. W., Song, J. B., Shu, X. X., Zhang, Y., and Yang, Z. M. (2013a). MiR395 is involved in detoxification of cadmium in Brassica napus. *J. Hazard. Mater.* 250–251, 204–211. doi:10.1016/j.jhazmat.2013.01.053.

Zhang, X., Lii, Y., Wu, Z., Polishko, A., Zhang, H., Chinnusamy, V., et al. (2013b). Mechanisms of small RNA generation from Cis-NATs in response to environmental and developmental cues. *Mol. Plant* 6, 704–715. doi:10.1093/mp/sst051.

Zhang, Z., Yu, J., Li, D., Zhang, Z., Liu, F., Zhou, X., et al. (2010). PMRD: plant microRNA database. *Nucleic Acids Res.* 38, D806-13. doi:10.1093/nar/gkp818.

Zhao, H., Wu, L., Chai, T., Zhang, Y., Tan, J., and Ma, S. (2012). The effects of copper, manganese and zinc on plant growth and elemental accumulation in the manganese-

hyperaccumulator Phytolacca americana. *J. Plant Physiol.* 169, 1243–52. doi:10.1016/j.jplph.2012.04.016.

Zhao, L., Sun, Y. Le, Cui, S. X., Chen, M., Yang, H. M., Liu, H. M., et al. (2011). Cd-induced changes in leaf proteome of the hyperaccumulator plant Phytolacca americana. *Chemosphere* 85, 56–66. doi:10.1016/j.chemosphere.2011.06.029.

Zhao, W., Li, Z., Fan, J., Hu, C., Yang, R., Qi, X., et al. (2015). Identification of jasmonic acid-associated microRNAs and characterization of the regulatory roles of the miR319/TCP4 module under root-knot nematode stress in tomato. *J. Exp. Bot.* 66, 4653–4667. doi:10.1093/jxb/erv238.

Zhuang, Y., and Tripp, E. A. (2017). The draft genome of Ruellia speciosa (Beautiful Wild Petunia: Acanthaceae). *DNA Res.* 24, 179–192. doi:10.1093/dnares/dsw054.

Zou, C., Sun, K., Mackaluso, J. D., Seddon, A. E., Jin, R., Thomashow, M. F., et al. (2011). Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. *Proc. Natl. Acad. Sci.* 108, 14992–14997. doi:10.1073/pnas.1103202108.

Zoubenko, O., Hudak, K., and Tumer, N. E. (2000). A non-toxic pokeweed antiviral protein mutant inhibits pathogen infection via a novel salicylic acid-independent pathway. *Plant Mol. Biol.* 44, 219–229.

Zoubenko, O., Uckun, F., Hur, Y., Chet, I., and Tumer, N. (1997). Plant resistance to fungal infection induced by nontoxic pokeweed antiviral protein mutants. *Nat. Biotechnol.* 15, 992–996. doi:10.1038/nbt1097-992.

# Appendices

**Appendix A: A small RNA targets pokeweed antiviral protein transcript**

Published manuscript

Klenov, A., **Neller, K. C. M.**, Burns, L. A., Krivdova, G., and Hudak, K. A. (2015). *Physiologia Plantarum* 156 (3), 241-51.

**Author contributions:** LB and KH conceived the study. AK and GK performed wet-lab experiments. KN performed bioinformatic analyses and drafted the manuscript. KH edited the manuscript.

Physiologia Plantarum
An International Journal for Plant Biology





# A small RNA targets pokeweed antiviral protein transcript

Alexander Klenov, Kira C. M. Neller, Lydia A. Burns, Gabriela Krivdova and Katalin A. Hudak*

Department of Biology, York University, Toronto, Ontario, Canada

**Correspondence**
*Corresponding author,
e-mail: hudak@yorku.ca

Received 29 July 2015;
revised 9 September 2015

doi:10.1111/ppl.12393

Ribosome-inactivating proteins (RIPs) are a class of plant defense proteins with N-glycosidase activity (EC 3.2.2.22). Pokeweed antiviral protein (PAP) is a Type I RIP isolated from the pokeweed plant, *Phytolacca americana*, thought to confer broad-spectrum virus resistance in this plant. Through a combination of standard molecular techniques and RNA sequencing analysis, we report here that a small RNA binds and cleaves the open reading frame of PAP mRNA. Additionally, sRNA targeting of PAP is dependent on jasmonic acid (JA), a plant hormone important for defense against pathogen infection and herbivory. Levels of small RNA increased with JA treatment, as did levels of PAP mRNA and protein, suggesting that the small RNA functions to moderate the expression of PAP in response to this hormone. The association between JA and PAP expression, mediated by sRNA299, situates PAP within a signaling pathway initiated by biotic stress. The consensus sequence of sRNA299 was obtained through bioinformatic analysis of pokeweed small RNA sequencing. To our knowledge, this is the first account of a sRNA targeting a RIP gene.

## Introduction

Plants have evolved numerous mechanisms to inhibit pathogen infection. Ribosome-inactivating proteins (RIPs) are a class of plant defense proteins with *N*-glycosidase activity. Specifically, RIPs catalyze the removal of purines from nucleic acid targets. Their first identified role was depurination of the conserved sarcin/ricin loop of large ribosomal RNA and a resultant inhibition of protein synthesis in vitro (Gessner and Irvin 1980, Endo et al. 1988). More recently, depurination of RNA-based viruses has been shown (Rajamohan et al. 1999, Karran and Hudak 2008, Kaur et al. 2011). Pokeweed antiviral protein (PAP), a Type I RIP isolated from the pokeweed plant *Phytolacca americana* (Irvin 1975) inhibits plant viruses including potato virus X and Y, tobacco mosaic virus and brome mosaic virus (Lodge et al. 1993, Karran and Hudak 2008). Transgenic plants expressing PAP exhibit virus resistance; however, these plants also display phenotypes indicative of toxicity, correlated with the level of PAP synthesis (Lodge et al. 1993). Unregulated toxicity of PAP is not observed in the native plant, pokeweed, suggesting the absence of important regulatory mechanisms in heterologous systems.

Little is known about the regulation of PAP expression in pokeweed, with the exception of preliminary temporal and spatial characterization (Honjo et al. 2002, Irvin et al. 1980, Kataoka et al. 1992). RIP induction by pathogen stress-related compounds, mechanical wounding, plant viruses and fungal pathogens has been reported (Dunaeva et al. 1999, Song et al. 2000, Qin et al. 2005, Iglesias et al. 2008). Temporal regulation of some RIPs has also been shown (Parente et al. 2008, Kawade and Masuda 2009, Loss-Morais et al. 2013). Although there has been some insight into transcriptional regulation of RIPs, their post-transcriptional regulation is not well characterized. Generally, positive correlation exists between transcript and protein levels (Kawade et al. 2008); however, discrepancies have been observed in some species

*Abbreviations* – EMSA, electrophoretic mobility shift assay; JA, jasmonic acid; LMW, low molecular weight; miRNA, micro RNA; nt, nucleotide; PAP, pokeweed antiviral protein; RACE, rapid extension of cDNA ends; RIP, ribosome-inactivating protein; siRNA, small interfering RNA; sRNA, small RNA.

(Iglesias et al. 2008,Vepachedu et al. 2003). Therefore, post-transcriptional regulation may be important for maintaining normal RIP expression patterns.

Small RNA (sRNA)-mediated regulation is a form of post-transcriptional control that has not been described for RIPs. sRNAs in plants are 20–24 nucleotide (nt), non-coding RNAs involved in fine-tuning gene expression. Depending on their biogenesis, sRNAs are either classified as microRNAs (miRNAs) or small interfering RNAs (siRNAs). Both types of sRNA participate in gene targeting mediated by the ARGONAUTE-containing RNA-Induced Silencing Complex (reviewed in Bologna and Voinnet 2014). In plants, sRNAs are induced in response to various biotic stresses, including pathogen infection, herbivory and related plant stress hormones (reviewed in Khraiwesh et al. 2012).

Based on the abundance of sRNAs involved in stress responses, and the fact that PAP is an antiviral RIP, we hypothesized that control of PAP expression involved sRNA-mediated gene regulation. Here, we identify a sRNA that causes cleavage of PAP mRNA, within its open reading frame. We also show that jasmonic acid (JA), a plant hormone important during herbivory and pathogen stress, differentially affects the expression of PAP mRNA and the sRNA. To our knowledge, this is the first report of a sRNA targeting a RIP message and provides novel insight into the post-transcriptional control of these plant defense enzymes.

## Materials and methods

### Pokeweed growth conditions and treatment with JA

Pokeweed seeds were immersed in 37% sulfuric acid for 5 min, rinsed in water and imbibed in water for 4 days at room temperature. Seeds were germinated in soil (Promix BX) and raised in a growth chamber (AC60, Biochambers, MB, CA) under fluorescent and incandescent lights at 180 $\mu mol\, m^{-2}\, s^{-1}$. Plants were fertilized once every 2 weeks with N:P:K 20:20:20. Plants at the 4-leaf stage were sprayed with 5 ml of 5 m*M* JA dissolved in 0.5% ethanol (to improve the solubility of JA). Negative control plants were sprayed with 0.5% ethanol alone. Plants were returned to the chamber and leaf tissue was harvested 24 h after treatment. All analyses for this study were conducted on pokeweed plants at the 4-leaf stage of growth.

### Isolation of total and low molecular weight RNA

Total RNA was extracted from pokeweed to probe for the sRNA, and cabbage served as the negative control. Leaf tissue (7 g) was ground to a powder with liquid nitrogen. TRIzol reagent (Molecular Research Center, Inc., Cincinnati, OH, USA) and chloroform were added, samples were vortexed and incubated at room temperature for 10 min. Particulate matter was pelleted by centrifugation at 16 000 *g* for 15 min at 4°C. Isopropanol was added to supernatants and the RNA was precipitated at −20°C. Samples were centrifuged at 16 000 g for 20 min at 4°C, RNA pellets were resuspended in water and extracted in acidic phenol:chloroform:isoamyl alcohol (25:24:1). RNA was precipitated in ethanol and finally resuspended in water and stored at −80°C.

LMW RNA was extracted from pokeweed and cabbage to serve as an enriched pool of sRNA. Following isopropanol precipitation of total RNA, the RNA was diluted to 1 $\mu g\, \mu l^{-1}$ in water and incubated in 5% PEG8000 and 0.5 *M* NaCl for 30 min at −20°C. Samples were centrifuged at 16 000 *g* for 25 min at 4°C to pellet the high molecular weight RNA. Supernatants were incubated in isopropanol at −20°C and LMW RNA was pelleted by centrifugation at 16 000 *g* for 25 min at 4°C, washed in 75% ethanol and resuspended in water.

### Northern blotting

Total pokeweed (PW) and cabbage (CB) RNA (38 µg) was separated through 1.5% agarose gel in 7% formamide, transferred to nylon membrane by capillary action and cross-linked to the membrane with UV light (120 $mJ\, cm^{-2}$ for 12 s). The RNA was probed by incubation with an $\alpha^{33}$P-CTP labeled riboprobe ($1 \times 10^6$ cpm) specific for a 245 nt portion of the PAP mRNA (Accession # AR009535.1). The internally labeled minus-strand RNA probe was synthesized by in vitro transcription with T7 polymerase from a polymerase chain reaction (PCR) template generated from a portion of the PAP mRNA spanning nt 680–925. The PAP-specific primers were PAP680for: 5′ gggagtaaaatcaagaagtcagg 3′ and PAP T7 925rev: 5′ taatacgactcactataggaaatcttaccccatgtctcttg 3′. An in vitro transcript of full-length PAP mRNA was used as a positive control.

### Primer extension

A PAP-specific reverse primer (5′ gaagatcattcggaaaagtgg 3′) complementary to nt 340–361 of PAP mRNA was 5′ end-labeled with $\gamma^{33}$P-ATP. Radiolabeled primer ($5 \times 10^5$ cpm) was denatured with total PW or CB RNA (30 µg) and annealed at room temperature for 15 min. The primer was extended by reverse transcription with 100 units of Superscript III (Life Technologies, Burlington, ON, CA) and incubation at 48°C for 90 min. Reactions were terminated by the addition of formamide buffer and denatured cDNA products were separated

by 7 *M* urea/6% acrylamide gel. To identify the cDNA 3′ ends, a dideoxynucleotide sequencing ladder of the PAP cDNA was generated with the same primer used for primer extension of total RNA.

### 5′ rapid extension of cDNA ends

Total RNA from pokeweed (36 µg) was used as template to generate PAP cDNA using a PAP-specific reverse primer. Briefly, total RNA was denatured with PAP reverse primer complementary to nt 607–632 (5′ ctaacacgagaattggcatttgggc 3′) and allowed to anneal at room temperature for 15 min. Reverse transcription was carried out at 48°C for 90 min using 100 units of Superscript III (Life Technologies, Burlington, ON, CA). The RNA was subsequently digested with 1.25 U of RNase H (New England Biolabs, Whitby, ON, CA) for 30 min at 37°C, and the cDNA was collected through Biobasic EZ-10 columns. cDNA samples (8.5 µl) were ligated with 10 units of T4 RNA Ligase 1 (New England Biolabs, Whitby, ON, CA) to 2.5 nmol of adapter primer (5′ $PO_4$- ccatggcaataccggtaaggtcctcactc 3′) in 1X RACE Buffer (20% PEG8000, 50 m*M* Tris–HCl pH 8.0, 10 m*M* $MgCl_2$, 10 ng µl$^{-1}$ BSA, 2 m*M* hexamine cobalt chloride, 20 µ*M* ATP) at 22°C for 4 h. After purification through an EZ-10 column, 5 µl of ligation product was amplified with two rounds of PCR with adapter-specific (AP1: 5′ gagtgaggaccttaccgg 3′, AP2: 5′ ccttaccggtattgccatgg 3′) and PAP-specific primers (PAP532-512R: 5′ caaagggatcagaataaccc 3′, PAP432-411BglIIR: 5′ agatcttggatttgtatttgtattggg 3′). The final PCR product was isolated from low-melt agarose and sequenced.

### Electrophoretic mobility shift assay

LMW RNA was probed with an $\alpha^{33}$P-CTP radiolabeled sense RNA transcript of the PAP sequence from nt 286–309. Complementary cDNA primers (PAP T7 286–302 for: 5′ taatacgactcactatagggctgtgaatacaatc 3′; PAP 309–293 rev: 5′ tgtagatgattgtattc 3′) were annealed, with the forward primer also containing a T7 polymerase-binding site. Subsequent PCR produced sufficient DNA template for in vitro transcription of the riboprobe. LMW RNA (10 µg) of pokeweed and $1.2 \times 10^5$ cpm of riboprobe were denatured at 95°C for 5 min in RNA-binding buffer (5 m*M* HEPES-KOH pH 7.8, 100 m*M* KCl, 6 m*M* $MgCl_2$, 3.8% glycerol). LMW RNA from cabbage (10 µg) was also incubated with riboprobe as a negative control. Samples were incubated at room temperature for 30 min and separated through a 15% non-denaturing acrylamide gel. Gels were dried under vacuum and bands were visualized with a phosphorimager. As a positive control for band shift, the riboprobe was incubated with a complementary anti-sense RNA, generated by in vitro transcription of a PCR product with T7 polymerase-binding site at the 3′ end of the template (PAP 286–302 for: 5′ ggctgtgaatacaatc 3′; PAP T7 309–293 rev: 5′ taatacgactcactatagggtgtagatgattg tattc 3′).

### Target construct for small RNA

The PAP cDNA with 2X FLAG sequence at the 5′ end was cloned downstream of the CaMV 35S promoter placed in the pBluescript vector. Primers were designed to introduce scrambled sequence at the putative sRNA299 target site within the PAP mRNA at nt 288–308. PCR using the PAP construct as template with primers NcoI2XFlagPAP216for (5′ gatgatccatggatggac tacaaagaccatgacggtgattataaagatcatgacatcgaagggaagatga agtcgatgcttgtggtg 3′) and Scram288-305PAP263-287rev (5′ tcgtcggtgcggacaataccaagttgaagttggtgcaagaatg 3′) generated the 5′ fragment of PAP, while PCR with primers BglIIstopPAPrev (5′ gatgatagatcttcaagttgtctgacag ctcccac 3′) and Scram291-308PAP309-331for (5′ tgtccg caccgacgagcaaatgttggaagtaccaccattag 3′) generated the 3′ fragment of PAP. Subsequent overlapping PCR using these two products as template, and primers NcoI2XFlagPAP216for and BglIIstopPAPrev, generated the full-length PAP construct with a scrambled sRNA299 target site. The scrambled sequence altered the amino acid sequence within this 21 nt section but did not introduce rare codons or stop codons that would cause ribosomes to stall and trigger mRNA decay. The wild-type PAP cDNA was amplified with the same PAP-specific 5′ and 3′ primers as used for the overlapping PCR.

### Isolation and transfection of pokeweed protoplasts

Protoplasts were isolated essentially as described (Koch et al. 1996) with minor modifications. Pokeweed leaves (3 g) were sliced into thin (2 mm) strips and incubated with enzyme solution (0.4 *M* mannitol, 20 m*M* MES pH 5.7, 1% cellulose RS-10, 0.15% macerozyme, 0.2% BSA) for 3 h. The mixture was filtered through cheesecloth and centrifuged in an IEC clinical centrifuge at speed 2 for 5 min. Pellets were resuspended in 0.55 *M* mannitol and overlayed onto 20% sucrose cushions. Following centrifugation in an IEC clinical centrifuge at speed 2 for 5 min, protoplasts were collected from the interface. Protoplasts were washed in W5 solution (154 m*M* NaCl, 125 m*M* $CaCl_2$, 5 m*M* KCl, 2 m*M* MES pH 5.7), incubated on ice for 30 min and centrifuged again at the same speed. Protoplast pellets were resuspended in MMg solution (400 m*M* mannitol,

15 m*M* $MgCl_2$, 4 m*M* MES pH 5.7) and counted with a hemocytometer.

Protoplast transfection was performed at room temperature. Wild-type and scrambled FLAG-tagged PAP constructs (30 μg) were mixed with 100 μl of protoplasts ($5 \times 10^4$ cells $ml^{-1}$) and 110 μl PEG/Ca solution (200 m*M* mannitol, 100 m*M* $CaCl_2$, 40% PEG4000, 50 m*M* PIPES pH 7.0) and incubated for 20 min. Samples were diluted with 0.44 ml of W5 solution followed by centrifugation in an IEC clinical centrifuge at speed 3 for 1 min. Pellets were resuspended in 100 μl of W5 solution and 1 ml of incubation medium (0.275 *M* mannitol, 1X Aoki salts, 10% sucrose, 1 μg $ml^{-1}$ gentamycin). Protoplasts were incubated under low light (25 μmol $m^{-2}$ $s^{-1}$) at room temperature for 20 h. After the incubation, samples were centrifuged at 1000 *g* for 10 min to recover the protoplasts for total RNA isolation.

### Reverse transcription-quantitative PCR

To quantify levels of PAP mRNA following JA treatment of plants, total PW RNA (8 μg) was reverse transcribed with a PAP-specific primer (PAP925Rev; 5′ gaaatcttaccccatgtctcttgc 3′) and 100 units of Superscript III (Life Technologies, Burlington, ON, CA) at 48°C for 90 min. The cDNA product (14.5 μl) was used as template for quantitative PCR using SYBR Green PCR Master Mix (Clontech, Mountain View, CA, USA) with PAP-specific forward (PAP680For; 5′ gggagtaaaatcagaagtcagg 3′) and reverse (PAP925Rev; 5′ gaaatcttaccccatgtctcttgc 3′) primers, in 20 μl total sample volume. To quantify levels of FLAG-tagged wild-type PAP or FLAG-tagged PAP scrambled mRNAs, total RNA from pokeweed protoplasts (20 μg) was reverse transcribed as above with a PAP-specific primer (PAP370-338rev; 5′ gcttcattac gaagatcattcagaaaagtggcg 3′). Quantitative PCR was performed as above with cDNA (14.5 μl), a FLAG-specific forward primer (2X FLAG; 5′ gactacaaagaccatgacggtgat tataaagatcatgac 3′) and the same primer used for cDNA synthesis. In both cases, the 28S rRNA was used as an internal control, with forward (5′ aacgtgagctgggtttagac cgtcg 3′) and reverse (5′ tcagtagggtaaaactaacc 3′) primers specific to 28S rRNA. To quantify levels of sRNA299 following JA treatment, we followed the method of stem-loop PCR essentially as described (Varkonyi-Gasic et al. 2007). LMW RNA (5 μg) was reverse transcribed as above with a sRNA299-specific hairpin primer (s299HP: 5′ gttggctctggtgcagtgtgagaggtatgcgcaccagagc caacacgtttgggc 3′). Quantitative PCR was performed as above with cDNA (14.5 μl), a sRNA299-specific forward primer (sRNA299For: 5′ gcggcggtgtagat-gattgtatt 3′) and a hairpin-specific universal primer (UniversalRev: 5′ acctctcacactgcac 3′). miR156 was used as an internal control as this miRNA has been validated as a suitable reference for biotic stress studies in other plants (Kulcheski et al. 2010). LMW RNA (5 μg) was reverse transcribed with a miR156-specific hairpin primer (miR156HP: 5′ gttggctctggtgcagtgtgagaggtatgcgcaccagagccaacacgt-gtgctc 3′) followed by quantitative PCR (qPCR) using miR156-specific forward primer (miR156For: 5′ gcggcg gtgacagaagagag 3′) and the hairpin-specific universal primer used above.

### Cellular lysate preparation and immunoblot

Three leaf discs per sample (1 cm diameter each) were taken from pokeweed plants and homogenized in Buffer A (25 m*M* Tris–HCl, pH 7.5, 1 m*M* EGTA, 1 m*M* DTT, 1 m*M* PMSF, 5% glycerol). Samples were clarified by centrifugation at 16 000 *g* for 5 min at 4°C. Supernatant protein was quantified using the Bradford Assay, and 5 μg protein per sample was separated through 12% SDS-PAGE. Proteins were transferred to nitrocellulose and probed with a polyclonal antibody specific for PAP (1:5000). Blots were also probed with a polyclonal antibody for L3, a ribosomal protein, as a loading control (1:5000). PAP and L3 were visualized by chemiluminescence.

### Small RNA sequencing

Total RNA from leaves of pokeweed plants treated with or without JA was sent to the Genomics Core Facility (Sunnybrook Hospital, Toronto, CA) for processing. In total, three pokeweed sRNA libraries were constructed from one (–JA) or two (+JA) plants using the strand-specific SOLiD Total RNA-Seq Kit (Life Technologies, Burlington, ON, CA). Briefly, preparation of the sRNA libraries was as follows: acrylamide gel purification of the 15–30 nt size fraction of RNA; simultaneous ligation of directional 5p and 3p primers; cDNA synthesis and acrylamide gel purification; PCR amplification (15 cycles) and column purification. Sequencing of the sRNA libraries was performed on a SOLiD 5500 XL machine. For each sRNA library, two technical replicates were sequenced.

### Bioinformatics analysis of the small RNA libraries

All low quality reads (containing a fastqcssanger value below 10) and those outside of the 18–28 nt size range were removed. Adapter sequences were trimmed with Clip v.1.0.1 from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). sRNA sequences derived from rRNA, tRNA, snRNA and snoRNA were identified with Bowtie v.1.1.0 (Langmead et al.

2009) and removed, using the *Beta vulgaris* genome (RefBeet-1.1) as a reference. *B. vulgaris* was chosen because it is the closest related species to pokeweed with a sequenced genome available (Dohm et al. 2014). Remaining reads in each library were grouped if they had the same sequence and termed 'unique sequences'. To identify potential regulatory sRNAs, libraries were aligned to the reverse complement strand of PAP mRNA (Accession # AR009535.1) with Bowtie v.1.1.0. Up to three mismatches were allowed in alignments. The Integrated Genomics Viewer v.2.3 (http://www.broadinstitute.org/igv/) was used to visualize and count aligned reads. A multiple sequence alignment was generated with Clustal Omega v.1.2.1 (http://www.ebi.ac.uk/Tools/msa/clustalo/) for sRNA299 reads in the –JA and +JA libraries.

### Statistical analysis

Significance testing was performed for all quantitative reverse-transcription PCR (RT-PCR) results. A one-tailed, unpaired Student's *t*-test was conducted using GraphPad Prism v.5.01.

## Results

### PAP mRNA is cleaved 298 nucleotides 3′ of the transcription start site

To investigate whether PAP mRNA may be targeted for cleavage by a sRNA, total RNA from pokeweed plants (PW) was probed for PAP mRNA by northern blot. Total cabbage (CB) RNA was used as a negative control, based on a literature search indicating no known RIP in this plant. We detected full-length PAP mRNA, based on comparison with an in vitro transcript of PAP, plus a lower molecular mass band, suggesting cleavage of PAP mRNA (Fig. 1A). To identify the site of cleavage, a gene-specific primer for PAP was extended by reverse transcriptase on total RNA from the same samples. A band indicating termination of extension was observed in pokeweed RNA but not in cabbage RNA (Fig. 1B). Sequencing of a PAP cDNA with the same reverse primer indicated that the cDNA terminated at nt A299, relative to the 5′ transcription start site of PAP mRNA. 5′ rapid extension of cDNA ends analysis was also conducted on total PW RNA. A PCR product indicative of full-length PAP mRNA was observed, along with a smaller PCR product, supporting truncation of the PAP mRNA (Fig. 1C). Sequencing of the smaller product confirmed that PAP mRNA was cleaved between nt C298 and A299. This cleavage occurred within the open reading frame of the PAP mRNA (Fig. 1D) and indicates the mRNA may be targeted by a small RNA.

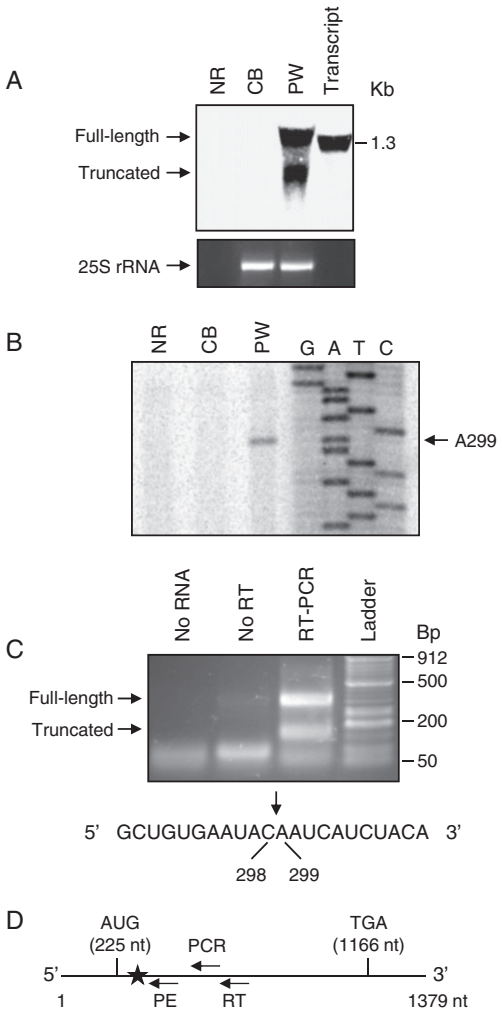



**Fig. 1.** PAP mRNA is cleaved 298 nucleotides 3′ of the transcription start site. (A) Total PW and CB RNA (38 μg) was separated through 1.5% agarose gel in 7% formamide, transferred to nylon membrane and probed with a PAP-specific minus-strand riboprobe. An in vitro transcript of full-length PAP mRNA (0.5 μg) was also separated as a positive control and a relative size marker. Prior to transfer, the gel was stained with ethidium bromide to visualize the 25S rRNA, as an indication of amount of total RNA loaded. (B) Pokeweed total RNA (PW) was template for primer extension using PAP-specific radiolabeled reverse primer complementary to region 340–361 nt. CB RNA was analyzed as a negative control, and No RNA (NR) indicates radiolabeled reverse primer without extension template. cDNA products were separated through 7 *M* urea/6% acrylamide gel and visualized with a phosphorimager. Dideoxynucleotide sequencing of a PAP cDNA with the same primer was used to identify the reverse transcriptase stall site. (C) 5′ RACE analysis of total PW RNA. PCR products were separated on 1% agarose gel and stained with ethidium bromide. Sequencing results of the cleaved PCR product indicate PAP mRNA cleavage between C298 and A299 (arrow) of full-length PAP mRNA. Both gels are representative of four independent experiments. (D) Diagram of full-length PAP mRNA with nucleotide number (1379 nt), indicating site of cleavage (star; 298–299 nt) and location of reverse primers used for primer extension (PE; 340–361 nt), reverse transcriptase (RT; 607–632 nt) and PCR (532–512 nt) for 5′ RACE analyses.

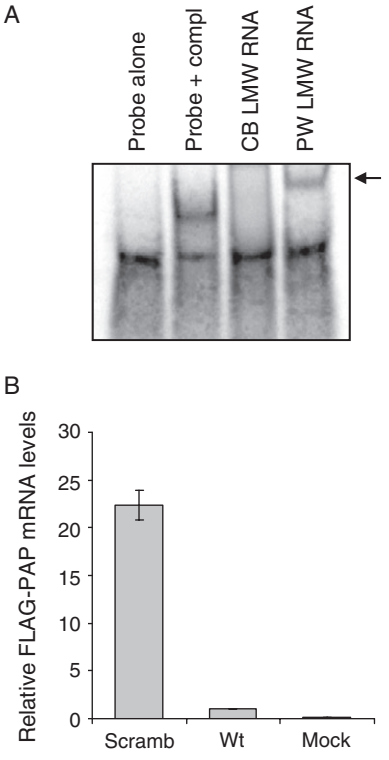



**Fig. 2.** sRNA299 targets PAP mRNA in vivo. (A) LMW RNA (10 μg) of PW and cabbage (CB) were incubated with a radiolabeled riboprobe of the PAP sequence from nt 286–309 downstream of the transcription start site. Following incubation, samples were separated on a non-denaturing, 15% acrylamide gel and visualized with a phosphorimager. As a positive control for band shift, the riboprobe was incubated with its complementary non-radiolabeled sequence (probe + complement). The gel is representative of three independent experiments. (B) Pokeweed protoplasts were transfected with either FLAG-tagged wild-type PAP cDNA or FLAG-tagged PAP cDNA with scrambled sRNA299 target sequence. Quantitative RT-PCR measurements of scrambled FLAG-PAP mRNA are indicated relative to wild-type FLAG-PAP mRNA. 28S rRNA was used as an internal control. Mock represents cells transfected with vector lacking insert. Values are means ± SE for five independent experiments.

## A small RNA binds to PAP mRNA in vivo

To test whether a sRNA from pokeweed bound to the observed cleavage site of PAP mRNA, we performed an RNA electrophoretic mobility shift assay. A (+)-strand, radiolabeled in vitro transcript of the 23 nt region of PAP mRNA potentially targeted by the sRNA was synthesized, based on the assumption that cleavage of PAP mRNA at A299 would result from the binding of a reverse complementary sRNA extending approximately 11 nt in the 5′ and 3′ direction from A299. This was reasoned, as plant miRNA-target interactions tend to cause cleavage between positions 10 and 11 of the alignment (German et al. 2008). The 23 nt transcript was used as a probe and incubated with the low molecular weight RNA pool isolated from pokeweed. A resulting shift in migration of the probe was detected (Fig. 2A). This band (indicated by arrow) migrated slightly slower than the migration of the positive control, which was an unlabeled 23 nt in vitro transcript complementary to the probe. Therefore, a sRNA from pokeweed binds to the PAP mRNA sequence and may be longer than 23 nt. We designated this sRNA 'sRNA299', referring to its induced cleavage between nt 298/299 of the PAP mRNA sequence. Incubation of the probe with LMW RNA from cabbage served as a negative control and did not produce a shift in probe migration, indicating some specificity of the probe for a sRNA from pokeweed.

To determine whether sRNA299 reduced levels of PAP mRNA in vivo, constructs of FLAG-tagged PAP, to distinguish them from endogenous PAP, were transfected into pokeweed protoplasts. The constructs contained either the wild-type target sequence for sRNA299, or scrambled target sequence. Quantitative RT-PCR demonstrated that FLAG-PAP mRNA levels bearing the scrambled target sequence had a significant, 22-fold increase compared with FLAG-PAP mRNA levels with wild-type sequence ($P < 0.01$, Student's *t*-test; Fig. 2B). Taken together, these results show that a sRNA binds a specific sequence within the open reading frame of PAP mRNA in vivo, resulting in reduced levels of full-length PAP mRNA.

## JA increases the level of PAP mRNA and sRNA299

To investigate the possible physiological role of the sRNA, we treated pokeweed leaves with JA and measured the levels of PAP mRNA and sRNA299. Given that JA signals a defense response to herbivory and pathogen attack, we hypothesized that it would increase the levels of PAP mRNA and resulting protein. This was indeed the case, as application of 5 m*M* JA increased PAP mRNA sevenfold and protein levels beyond control plants without the hormone treatment (Fig. 3A, B). The increase in PAP mRNA with JA, relative to water, was significant ($P < 0.05$, Student's *t*-test). We initially compared the effect of water and 0.5% ethanol on the level of PAP mRNA because JA was prepared in 0.5% ethanol. However, there was no obvious difference between water and 0.5% ethanol on PAP mRNAs levels; therefore, subsequent analyses were performed on plants treated with JA in 0.5% ethanol (+JA) and compared with 0.5% ethanol alone (−JA). We quantified the level of sRNA299 under the same conditions and observed a significant, 3.4-fold increase with JA ($P < 0.01$, Student's *t*-test; Fig. 4A). Treatment of pokeweed plants with JA also increased the amount of PAP mRNA cleavage at the A299 target site, relative to untreated plants (Fig. 4B). These results suggest that PAP levels increase during JA stress, which agrees

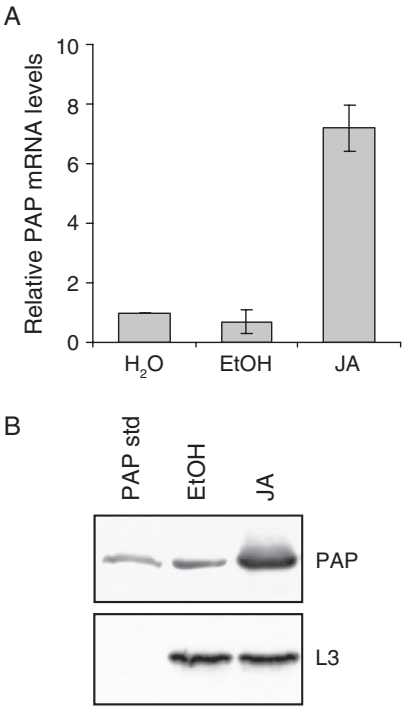



**Fig. 3.** JA treatment increases levels of PAP mRNA and protein. Pokeweed plants were treated with 5 m*M* JA 24 h prior to RNA or protein extraction. (A) Quantitative RT-PCR measurements of PAP mRNA from pokeweed total RNA are indicated. 28S rRNA was used as an internal control. Values are relative to the water-treated sample and are means ± SE for three independent experiments. (B) Total PW protein (5 μg) was separated by 12% SDS-PAGE, transferred to nitrocellulose and probed with a polyclonal antibody specific to PAP (1:5000). The blot was also probed with a polyclonal antibody specific for L3, a large ribosomal subunit protein, as a loading control (1:5000). PAP standard (std) is purified PAP from pokeweed. The blot is representative of three independent experiments.

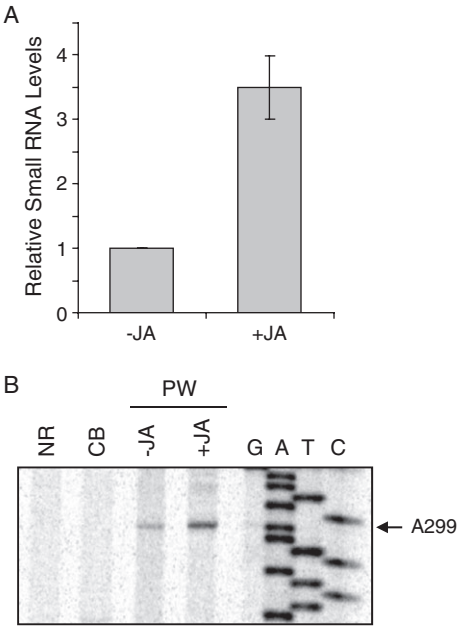



**Fig. 4.** JA treatment increases levels of sRNA299 and PAP mRNA cleavage. Pokeweed plants were treated with 5 m*M* JA 24 h prior to RNA extraction. (A) Quantitative RT-PCR measurements of sRNA299 from pokeweed low molecular weight RNA are indicated. Measurements used miR156 as an internal control. Values are relative to −JA and are means ± SE for three independent experiments. (B) Pokeweed total RNA (PW) was template for primer extension using PAP-specific radiolabeled reverse primer complementary to region 340–361 nt. CB RNA was analyzed as a negative control, and No RNA (NR) indicates radiolabeled reverse primer without extension template. cDNA products were separated through 7 *M* urea/6% acrylamide gel and visualized with a phosphorimager. Dideoxynucleotide sequencing of a PAP cDNA with the same primer was used to identify the reverse transcriptase stall site. The gel is representative of four independent experiments.

with the hypothesized role of PAP as a defense protein. Moreover, the increase in the sRNA suggests that it offsets a JA-induced transcriptional increase in PAP mRNA.

## Pokeweed shows a typical distribution of small RNA sequences

To characterize the sRNA content of pokeweed, we performed sRNA sequencing of plants treated with or without JA. The sRNA libraries illustrated a typical angiosperm-specific size distribution, with 23–24 nt RNAs being most abundant, followed by 20–22 nt RNAs (Fig. S1, Supporting Information). The distributions of unique sequences and total reads were similar. Averaging the results across unique sequences, reads and treatments, the percentages of 20 nt, 21 nt, 22 nt, 23 nt and 24 nt sRNAs were 2.29, 5.50, 5.30, 20.91 and 60.65, respectively. Because of the limited biological variability represented by our RNA-Seq samples, we did not investigate the effect of JA, if any, on the sRNA distribution in pokeweed.

## RNA-Seq confirms the existence of a small RNA targeting PAP mRNA

Sequencing of the small RNA pool from pokeweed revealed several small RNAs that aligned to the PAP mRNA, in both the plus- and minus-sense orientation. Based on the observed cleavage at ~300 nt of pokeweed mRNA, we expected to detect small RNAs aligning to this particular region. To identify sRNA299 in pokeweed, libraries were aligned to the reverse complement sequence of PAP mRNA. Between the –JA and +JA libraries, eight reads aligned to the previously characterized PAP mRNA target site. As determined by multiple sequence alignment, the reads ranged in length from 18

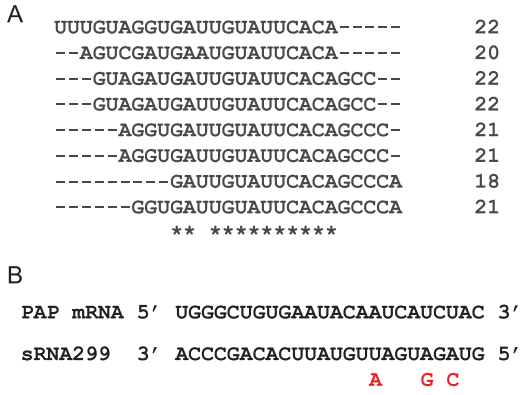



**Fig. 5.** RNA-Seq confirms the existence of sRNA299 in pokeweed. The +JA and −JA libraries were aligned to the PAP mRNA sequence to identify sRNAs that could regulate the transcript. (A) The multiple sequence alignment for sRNA299 reads present in the +JA and −JA libraries. (B) An illustration of sRNA299 binding to PAP mRNA via complementary base-pairing. Letters in red indicate nucleotides that are not complementary to PAP mRNA but were conserved in some sRNA299 reads.

to 22 nt and were highly conserved (Fig. 5A). The resultant sRNA299 consensus sequence showed high complementarity to the PAP mRNA target site (Fig. 5B). Of the eight reads, three were completely complementary to PAP mRNA and five deviated by only 1–3 nt. These differences could arise from sequencing errors or from different small RNA precursors. Taken together, these results confirm the presence of sRNA299 and its ability to target the PAP mRNA through complementary base-pairing.

## Discussion

We show here that PAP mRNA is targeted by a sRNA, which results in cleavage of the message within its open reading frame. Therefore, the sRNA negatively regulates PAP mRNA. Interestingly, JA treatment increases the levels of both PAP mRNA and the sRNA. In most documented cases, the level of sRNA is inversely related to the level of its target RNA; however, recent publications of genome-wide analyses demonstrate that several plant sRNA-target pairs have positively correlated expression patterns (He et al. 2013, Peng et al. 2014). Indeed, an equal number of negatively and positively correlated miRNA-target mRNA pairs have been described during tomato fruit development (Lopez-Gomollon et al. 2012). Positive correlations suggest mechanisms of action that are not limited to classic temporal regulation where negative correlation is evident. For example, miR395 and its target mRNA SULTR2;1 are both induced by low sulfur levels but each is transcribed in a separate neighboring cell type (Kawashima et al. 2009). Therefore, there is an exclusion between the small RNA and its target mRNA, and miR395 may function as a secondary means of preventing SULTR2:1 expression in specific cell types. Other scenarios describe how the mRNA is transcribed equally in two domains; yet, the small RNA is transcribed in only one, with the purpose of excluding the expression of the target in one domain only (Kidner and Martienssen 2004; Levine et al. 2007). Another type of regulation occurs when the small RNA diminishes or modulates the expression of the mRNA target and the degree of target reduction depends on the level of miRNA (Nikovics et al. 2006). Such a mechanism could play an important role in controlling the toxicity of PAP. We suggest that the increase in sRNA may be a means to limit the over-expression of PAP during induction. For example, transgenic plants expressing PAP from a constitutive 35S-CaMV promoter displayed symptoms of toxicity, including stunted growth, mottled leaves and sterility (Lodge et al. 1993). Authors reported correct localization of PAP, indicating deregulated PAP expression likely contributed to the observed toxicity.

sRNAs represent only one level of regulation of gene expression, which is also controlled by transcription factors/promoter sequences and mRNA stability. The contribution of each may mask the negative effect of a small RNA on target mRNA levels. RIP induction by hormones associated with biotic stress has been reported previously. For example, RIPs induced by plant hormones include JIP60, a JA-induced Type III RIP from barley (Dunaeva et al. 1999), beetin, a salicylic acid and hydrogen peroxide-induced Type I RIP from sugar beet (Iglesias et al. 2008) and soRIP2, a salicylic acid-induced Type I RIP from spinach (Kawade and Masuda 2009). Some RIPs are induced directly by stimuli upstream of the hormone signal, providing insight into their functional targets. For example, RIP2, a Type III RIP from maize with insecticidal activity, was induced 100-fold at the RNA level, with a concomitant increase in protein, upon caterpillar feeding but not mechanical wounding (Chuang et al. 2014). Other RIPs have shown pathogen-inducible expression patterns, including beetin and curcin, which are virus and fungus inducible, respectively (Girbés et al. 1996, Qin et al. 2010). Although the PAP gene promoter remains unknown, it is probable that JA treatment induces transcriptional activation of the PAP gene. Our finding that PAP expression is enhanced by JA agrees with previous reports which place RIPs within a signaling cascade mediated by biotic stress. While the antiviral activity of PAP is well established, the induction of PAP by JA suggests that PAP could play a more broad-spectrum role against herbivores and necrotrophic pathogens, well-known elicitors of JA signaling. The targeting of PAP by a small RNA that is also induced by JA may serve to fine-tune PAP levels during stress.

Our finding that PAP mRNA is targeted by a sRNA may explain some discrepancies reported previously between RIP transcript and protein levels. For example, in sugar beet, beetin transcripts were found at similar levels throughout plant development but the protein was only synthesized in adult plants (Iglesias et al. 2008). Upon treatment with mediators of systemic acquired resistance, beetin expression increased at the RNA and protein levels but only in adult plants. Authors suggested that variations in transcript stability, polyadenylation or protein factor recognition could account for developmental regulation of this Type I RIP. In this study, we found 5′ truncated PAP transcripts in pokeweed because of sRNA-associated cleavage. A developmentally regulated sRNA that targets beetin may also contribute to its observed expression pattern in sugar beet.

Although the abundance of sRNA299 in our RNA-Seq libraries was low, averaging less than 1 read per million (data not shown), it is possible that this sRNA shows a developmental or spatial expression pattern that is important for PAP regulation in pokeweed. A recent report found that the vast majority of miRNAs in plants are of low abundance and are species-specific (Montes et al. 2014). In contrast, highly abundant miRNA families tend to be well conserved among plants (Montes et al. 2014). RIPs have been found in less than 20% of angiosperm taxonomic orders, indicating that they are poorly conserved (Di Maro et al. 2014). We searched current plant sRNA databases (plant miRNA database: http://bioinformatics.cau.edu.cn/PMRD/ and tasiRNA database: http://bioinfo.jit.edu.cn/tasiRNADatabase/) for sequences aligning to sRNA299 and did not find any, suggesting that sRNA299 may be unique to pokeweed. Taken together, the low abundance and apparent lack of conservation of sRNA299 reported here does not undermine its functional importance in pokeweed.

Future work will focus on identifying the origin of sRNA299. Presently, our analysis selected for sRNA299 sequences with high complementarity to PAP mRNA and their presence suggests processing from a siRNA pathway. Our electrophoretic mobility shift assay results indicate that the sRNA299 probe annealed to a sRNA larger than 23 nt, although this could be attributed to the use of a non-denaturing gel, or the binding of a third component within the cellular lysate. To our knowledge, this is the first report of a sRNA targeting a RIP mRNA; the expression of other plant RIPs may be controlled in a similar manner. Additionally, this work suggests an endogenous mechanism to modulate the expression of a RIP, possibly limiting its associated toxicity. This may have important implications in transgenic plant systems and agricultural selection programs.

## Authors' contributions

L. A. B., K. C. M. N. and K. A. H. conceived the experiments. A. K., K. C. M. N. and G. K. designed and performed the experiments. K. C. M. N. drafted the manuscript. K. A. H. edited the manuscript. All authors approved the final version.

*Acknowledgements* – The authors are grateful to Dr Yutaka Amemiya, Manager, Genomics Core Facility, Next Generation Sequencing Lab, Sunnybrook Research Institute for small RNA sequencing and for his assistance and advice with sequence analysis. This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to K. A. H. and Canada Graduate Scholarships to K. C. M. N. and G. K.

## References

Bologna NG, Voinnet O (2014) The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. Annu Rev Plant Biol 65: 473–503

Chuang W-P, Herde M, Ray S, Castano-Duque L, Howe GA, Luthe DS (2014) Caterpillar attack triggers accumulation of the toxic maize protein RIP2. New Phytol 201: 928–939

Di Maro A, Citores L, Russo R, Iglesias R, Ferreras JM (2014) Sequence comparison and phylogenetic analysis by the maximum likelihood method of ribosome-inactivating proteins from angiosperms. Plant Mol Biol 85: 575–588

Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, Rupp O, Sörensen TR, Stracke R, Reinhardt R, Goesmann A, Kraft T, Schulz B, Stadler PF, Schmidt T, Galadón T, Lehrach H, Weisshaar B, Himmelbauer H (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505: 546–549

Dunaeva M, Goebel C, Wasternack C, Parthier B, Goerschen E (1999) The jasmonate-induced 60 kDa protein of barley exhibits N-glycosidase activity in vivo. FEBS Lett 452: 263–266

Endo Y, Tsurugi K, Lambert JM (1988) The site of action of six different ribosome-inactivating proteins from plants on eukaryotic ribosomes: the RNA N-glycosidase activity of the proteins. Biochem Biophys Res Commun 150: 1032–1036

German MA, Pillay M, Jeong DH, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, De Paoli E, Lu C, Schroth G, Meyers BC, Green PJ (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol 26: 941–946

Gessner SL, Irvin JD (1980) Inhibition of elongation factor 2-dependent translocation by the pokeweed antiviral protein and ricin. J Biol Chem 255: 3251–3253

Girbés T, de Torre C, Iglesias R, Ferrerras JM (1996) RIP for viruses. Nature 379: 777–778

He C, Cui K, Zhang J, Duan A, Zeng Y (2013) Next-generation sequencing-based mRNA and microRNA expression profiling analysis revealed pathways involved in the rapid growth of developing culms in Moso bamboo. BMC Plant Biol 13: 119

Honjo E, Dong D, Motoshima H, Watanabe K (2002) Genomic clones encoding two isoforms of pokeweed antiviral protein in seeds (PAP-S1 and S2) and the N-glycosidase activities of their recombinant proteins on ribosomes and DNA in comparison with other isoforms. J Biochem 131: 225–231

Iglesias R, Pérez Y, Citores L, Ferreras JM, Méndez E, Girbés T (2008) Elicitor-dependent expression of the ribosome-inactivating protein beetin is developmentally regulated. J Exp Bot 59: 1215–1223

Irvin J (1975) Purification and partial characterization of the antiviral protein from *Phytolacca americana* which inhibits eukaryotic protein synthesis. Arch Biochem Biophys 169: 522–528

Irvin JD, Kelly T, Robertus JD (1980) Purification and properties of a second antiviral protein from *Phytolacca americana* which inactivates eukaryotic ribosomes. Arch Biochem Biophys 200: 418–425

Karran RA, Hudak KA (2008) Depurination within the intergenic region of Brome mosaic virus RNA3 inhibits viral replication in vitro and in vivo. Nucleic Acids Res 36: 7230–7239

Kataoka J, Habuka N, Masuta C, Miyano M, Koiwai A (1992) Isolation and analysis of a genomic clone encoding a pokeweed antiviral protein. Plant Mol Biol 20: 879–886

Kaur I, Gupta RC, Puri M (2011) Ribosome inactivating proteins from plants inhibiting viruses. Virol Sin 26: 357–365

Kawade K, Masuda K (2009) Transcriptional control of two ribosome-inactivating protein genes expressed in spinach (*Spinacia oleracea*) embryos. Plant Physiol Biochem 47: 327–334

Kawade K, Ishizaki T, Masuda K (2008) Differential expression of ribosome-inactivating protein genes during somatic embryogenesis in spinach (Spinacia oleracea). Physiol Plant 134: 270–281

Kawashima CG, Yoshimoto N, Maruyama-Nakashita A, Tsuchiya YN, Saito K, Takahashi H, Dalmay T (2009) Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. Plant J 57: 313–321

Khraiwesh B, Zhu J-K, Zhu J (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim Biophys Acta 1819: 137–148

Kidner CA, Martienssen RA (2004) Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature 428: 81–84

Koch PE, Bonness MS, Lu H, Mabry TJ (1996) Protoplasts from *Phytolacca dodecandra* L'Herit (endod) and *P. americana* L. (pokeweed). Plant Cell Rep 15: 824–828

Kulcheski FR, Marcelino-Guimaraes FC, Nepomuceno AL, Abdelnoor RV, Margis R (2010) The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. Anal Biochem 406: 185–192

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25

Levine E, McHale P, Levine H (2007) Small regulatory RNAs may sharpen spatial expression patterns. PLoS Comput Biol 3: e233

Lodge JK, Kaniewski WK, Tumer NE (1993) Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. Proc Natl Acad Sci USA 90: 7089–7093

Lopez-Gomollon S, Mohorianu I, Szittya G, Moulton V, Dalmay T (2012) Diverse correlation patterns between microRNAs and their targets during tomato fruit development indicates different modes of microRNA actions. Planta 236: 1875–1887

Loss-Morais G, Turchetto-Zolet AC, Etges M, Cagliari A, Körbes AP, Maraschin Fdos S, Margis-Pinheiro M, Margis R (2013) Analysis of castor bean ribosome-inactivating proteins and their gene expression during seed development. Genet Mol Biol 36: 74–86

Montes RA, de Fátima R-CF, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, Marsch-Martínez N, Meyers BC, Green PJ, de Folter S (2014) Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun 5: 3722

Nikovics K, Blein T, Peaucelle A, Ishida T, Morin H, Aida M, Laufs P (2006) The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. Plant Cell 18: 2929–2945

Parente A, Conforto B, Di Maro A, Chambery A, De Luca P, Bolognesi A, Iriti M, Faoro F (2008) Type 1 ribosome-inactivating proteins from *Phytolacca dioica* L. leaves: differential seasonal and age expression, and cellular localization. Planta 228: 963–975

Peng Z, He S, Gong W, Sun J, Pan Z, Xu F, Lu Y, Du X (2014) Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. BMC Genomics 15: 760

Qin W, Ming-Xing H, Ying X, Xin-Shen Z, Fang C (2005) Expression of a ribosome inactivating protein (curcin 2) in *Jatropha curcas* is induced by stress. J Biosci 30: 351–357

Qin X, Shao C, Hou P, Gao J, Lei N, Jiang L, Ye S, Gou C, Luo S, Zheng X, Gu X, Zhu X, Xu Y, Chen F (2010) Different functions and expression profiles of curcin and

curcin-L in *Jatropha curcas* L. Z Naturforsch C 65: 355–362
Rajamohan F, Venkatachalam TK, Irvin JD, Uckun FM (1999) Pokeweed antiviral protein isoforms PAP-I, PAP-II, and PAP-III depurinate RNA of human immunodeficiency virus (HIV)-1. Biochem Biophys Res Commun 260: 453–458
Song SK, Choi Y, Moon YH, Kim S-G, Choi YD, Lee JS (2000) Systemic induction of a *Phytolacca insularis* antiviral protein gene by mechanical wounding, jasmonic acid, and abscisic acid. Plant Mol Biol 43: 439–450
Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3: 12
Vepachedu R, Bais HP, Vivanco JM (2003) Molecular characterization and post-transcriptional regulation of ME1, a type-I ribosome-inactivating protein from *Mirabilis expansa*. Planta 217: 498–506

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Fig. S1.** Pokeweed shows a typical small RNA size distribution. RNA-Seq was performed from low molecular weight RNA extracted 24 h after treatment of pokeweed plants with or without 5 m*M* JA. (A) Size distribution of unique sequences. (B) Size distribution of reads. Values are means ± SE for two technical replicates from one plant (–JA) or two technical replicates from two plants (+JA).

Edited by H. Saitoh

Appendix B

## JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS



Aug 23, 2019

This Agreement between Prof. Katalin Hudak ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

| | |
|---|---|
| License Number | 4654810343282 |
| License date | Aug 23, 2019 |
| Licensed Content Publisher | John Wiley and Sons |
| Licensed Content Publication | Physiologia Plantarum |
| Licensed Content Title | A small RNA targets pokeweed antiviral protein transcript |
| Licensed Content Author | Katalin A. Hudak, Gabriela Krivdova, Lydia A. Burns, et al |
| Licensed Content Date | Nov 16, 2015 |
| Licensed Content Volume | 156 |
| Licensed Content Issue | 3 |
| Licensed Content Pages | 11 |
| Type of use | Dissertation/Thesis |
| Requestor type | Author of this Wiley article |
| Format | Print and electronic |
| Portion | Full article |
| Will you be translating? | No |
| Title of your thesis / dissertation | Characterization of defense-associated gene expression in Phytolacca americana using novel transcriptomic and genomic resources |
| Expected completion date | Aug 2019 |
| Expected size (number of pages) | 200 |
| Requestor Location | Prof. Katalin Hudak<br>Dept of Biology<br>York University<br>4700 Keele St.<br>Toronto, ON M3J 1P3<br>Canada<br>Attn: Prof. Katalin Hudak |
| Publisher Tax ID | EU826007151 |
| Total | 0.00 CAD |
| Terms and Conditions | |

### TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

**Terms and Conditions**

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley

Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition

of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

**WILEY OPEN ACCESS TERMS AND CONDITIONS**

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

**The Creative Commons Attribution License**

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

**Creative Commons Attribution Non-Commercial License**

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

**Creative Commons Attribution-Non-Commercial-NoDerivs License**

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

**Use by commercial "for-profit" organizations**

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library http://olabout.wiley.com/WileyCDA/Section/id-410895.html

**Other Terms and Conditions:**

**Appendix C: List of additional contributions**

Jobst, K. A., Klenov, A., **Neller, K. C. M.**, and Hudak, K. A. (2016). Effect of depurination on cellular and viral RNA. In: *RNA Technologies: Modified Nucleic Acids in Biology and Medicine.* J. Barciszewski, V.A. Erdmann and S. Jurga eds., Springer Publishing Co., Inc., New York, NY. ISBN 978-3-319-34175-0. pp 273-297.

Krivdova, G., **Neller, K. C. M.**, Parikh, B., and Hudak, K. A. (2014). Antiviral and antifungal properties of ribosome inactivating proteins. In: *Ribosome Inactivating Proteins: Biology and Applications.* D.A. Lappi and F. Stirpe eds., John Wiley & Sons, Inc., New Jersey, USA. ISBN-10: 1118125657; ISBN-13: 978-1118125656. pp 198-211.

**Neller, K. C. M.**, Klenov, A., and Hudak, K. A. (2019). Prediction and characterization of miRNA/target pairs in non-model plants using RNA-seq. *Curr. Protoc. Plant Biol.*, e20090. doi:10.1002/cppb.20090.